frontiers | Frontiers in Pharmacology

TYPE Review
PUBLISHED 20 August 2025
DOI 10.3389/fphar.2025.1609079

OPEN ACCESS

EDITED BY
Jong-Min Kim,
University of Minnesota Morris, United States

REVIEWED BY
Luigi Marano,
Academy of Applied Medical and Social Sciences, Poland
Zhao Yuxin,
Zhejiang Normal University, China

*CORRESPONDENCE
Weilong Zhao,
weilong.zhao@abbvie.com
Si Wu,
ws0113sleepingtree@gmail.com

†PRESENT ADDRESS
Si Wu,
Amgen, South San Francisco, CA, United States

RECEIVED 09 April 2025
ACCEPTED 22 July 2025
PUBLISHED 20 August 2025

CITATION
Zhang B, Wan Z, Luo Y, Zhao X, Samayoa J, Zhao W and Wu S (2025) Multimodal integration strategies for clinical application in oncology. *Front. Pharmacol.* 16:1609079. doi: 10.3389/fphar.2025.1609079

Multimodal integration strategies for clinical application in oncology

Baoyi Zhang[1], Zhuoya Wan[2], Yige Luo[1], Xi Zhao[1], Josue Samayoa[1], Weilong Zhao[1]* and Si Wu[1]*†

[1]AbbVie Bay Area, South San Francisco, CA, United States, [2]AbbVie, Inc., North Chicago, IL, United States

In clinical practice, a variety of techniques are employed to generate diverse data types for each cancer patient. These data types, spanning clinical, genomics, imaging, and other modalities, exhibit significant differences and possess distinct data structures. Therefore, most current analyses focus on a single data modality, limiting the potential of fully utilizing all available data and providing comprehensive insights. Artificial intelligence (AI) methods, adept at handling complex data structures, offer a powerful approach to efficiently integrate multimodal data. The insights derived from such models may ultimately expedite advancements in patient diagnosis, prognosis, and treatment responses. Here, we provide an overview of current advanced multimodal integration strategies and the related clinical potential in oncology field. We start from the key processing methods for single data modalities such as multi-omics, imaging data, and clinical notes. We then include diverse AI methods, covering traditional machine learning, representation learning, and vision language model, tailored to each distinct data modality. We further elaborate on popular multimodal integration strategies and discuss the related strength and weakness. Finally, we explore potential clinical applications including early detection/diagnosis, biomarker discovery, and prediction of clinical outcome. Additionally, we discuss ongoing challenges and outline potential future directions in the field.

KEYWORDS

deep learning, multimodal integration, oncology, prognosis, biomarker, treatment response

1 Introduction

The rapid advancement of high-throughput technologies (Mardis, 2019), coupled with the digitalization of healthcare and electronic health records (EHRs) adoption (Jensen et al., 2012), has led to an unprecedented explosion of multi-modal datasets in oncology. These diverse data modalities include, but are not limited to, patient clinical records, multi-omics data—spanning genomics, transcriptomics, proteomics, and metabolomics—at bulk, single-cell, and spatial levels, as well as medical imaging (magnetic resonance imaging [MRI], computed tomography [CT], histopathology) and wearable sensor data. Each of these modalities provides unique insights into cancer diagnosis (Carrillo-Perez et al., 2022; Cui C. et al., 2022), prognosis (Lobato-Delgado et al., 2022; Zhu et al., 2020), and treatment responses (Chen Z. et al., 2024; Keyl et al., 2025; Wang et al., 2023), yet their true potential lies in their integration (Boehm et al., 2022; Unger and Kather, 2024). Multi-modal data fusion enables the combination of orthogonal information, allowing different data types to

complement one another and augment the overall information content beyond what a single modality can provide (Kline et al., 2022; Miotto et al., 2018). By integrating these diverse datasets, researchers can achieve a more comprehensive understanding of complex biological processes, improve inference accuracy, and enhance clinical decision-making, ultimately driving advancements in precision oncology.

Although multi-modal integration holds great promise for improving disease modeling and biomarker prediction, it presents several challenges due to the scale, complexity, and heterogeneity of the data. A primary challenge is data heterogeneity (Li et al., 2022a; Tang Z. et al., 2024; Zhang et al., 2025), as different modalities often vary in format, structure, and coding standards, and may originate from multiple vendors or institutions, making normalization and harmonization crucial before integration. Additionally, data quality and completeness issues, such as missing values, inconsistencies, and noise, can compromise both integration efforts and model performance (Cui et al., 2022; Waqas et al., 2024; Zhao et al., 2024). The computational and storage demands of large-scale multi-modal datasets—particularly high-resolution imaging and raw genomics data—necessitate advanced infrastructure and scalable analytical tools to enable efficient data (pre)processing and integration. Furthermore, multi-modal fusion methods have evolved in diverse directions, yet standardized methodologies and workflows remain underdeveloped (Chen et al., 2023; Chen et al., 2021a). Addressing these challenges requires comprehensive data frameworks encompassing preprocessing, alignment, harmonization, and integration, along with improved storage solutions, computational resources, and interdisciplinary collaboration. Overcoming these barriers is key to unlocking the full potential of multi-modal data for precision medicine and clinical decision-making.

Over the past decades, artificial intelligence (AI) technologies have grown rapidly and demonstrated immense potential in clinically relevant tasks. They excel at handling complex datasets and extracting meaningful clinical insights—capabilities that typically require years of human training and experience. For instance, image-based models have been developed to assist in cancer diagnosis, staging, grading and subtyping by analyzing morphological features in histopathological images (Araújo et al., 2017; Arvaniti et al., 2018; Coudray et al., 2018; Kanavati et al., 2021; Nagpal et al., 2020; Wan et al., 2017). Large language models have been applied to transform unstructured clinical notes into structured data, facilitating centralized data strategy and enabling more efficient downstream analysis (Luo et al., 2022; Sorin et al., 2024; Van Veen et al., 2024). Moreover, AI is particularly well-suited for integrating diverse data modalities. Various deep learning models have been developed to infer genomic data from imaging data (Diao et al., 2021; El et al., 2024; Jané et al., 2023; Jin et al., 2024; Kather et al., 2020; Shamai et al., 2022; Zhang et al., 2024a). Therefore, AI-driven multimodal data integration can benefit multiple aspects of clinical development, including biomarker discovery, patient stratification, and clinical trial recruitment.

Although several reviews on AI methods for multimodal integration have been published in past years (Boehm et al., 2022; Waqas et al., 2024; Kline et al., 2022; Stahlschmidt et al., 2022), most have not detailed the technical aspects underlining these strategies. Additionally, discussions on emerging technologies such as spatial transcriptomics, single-cell sequencing, and most-recent advanced computational methods were not discussed due to the publication dates of those reviews. Thus, this review explores the technical specifics of both established and newly developed methods, while incorporating discussions on novel data types. We specifically highlight key AI approaches for integrating common data modalities and explore their potential applications in clinically relevant tasks. In the last section, we discuss existing challenges and future directions for advancing AI-driven multimodal data integration.

2 Unimodal processing

2.1 Omics

2.1.1 Traditional omics data

In the context of traditional omics data, unimodal processing refers to the independent analysis of a single data type before any cross-modality integration. This approach ensures rigorous quality control, effective noise reduction, and precise feature extraction. For instance, genomic analysis employs tools such as GATK (Valle et al., 2016), MuTect (Aaltonen et al., 2020), and VarScan (Koboldt et al., 2012) to detect mutations and structural variants—as demonstrated in the TCGA pan-cancer study that utilized whole-genome sequencing to identify critical mutations across various cancer types. Transcriptomic analysis uses methods like DESeq2 and EdgeR to quantify gene expression and determine differential expression patterns associated with distinct disease subtypes (Siavoshi et al., 2022; Varet et al., 2016).

To further simplify the complexity of high-dimensional gene expression data, dimension reduction techniques such as principal component analysis (PCA) capture the primary axes of variation, effectively summarizing the most variable transcriptional patterns across samples (Alter et al., 2000). These components can simplify complex datasets and have been used to identify molecular subtypes of cancer with distinct biological behaviors and prognostic implications (Perou et al., 2000). Additionally, gene signatures are predefined sets of genes whose collective expression patterns are associated with specific biological states or clinical phenotypes (Subramanian et al., 2005). They have been extensively used in current clinical workflow. For example, the PAM50 gene signature is widely used for classifying breast cancer into intrinsic subtypes with prognostic relevance (Parker et al., 2023). Systematic identification of gene signatures can improve patient stratification by correlating molecular features to disease progression and treatment response, as demonstrated by the use of Oncotype DX to assess breast cancer recurrence risk (Liberzon et al., 2015). Pathway-based approaches involve mapping significant gene expression changes onto known biological pathways, such as those cataloged in KEGG or Reactome databases (Kanehisa and Goto, 2000). These methods can reveal dysregulated cellular processes and signaling networks in cancer, providing functional insights into disease mechanisms and potential therapeutic targets, exemplified by a colorectal cancer study using KEGG pathway enrichment analysis (Croft et al., 2011). Pathway analysis has been especially useful for interpreting high-dimensional transcriptomic data and linking molecular alterations to broader biological functions (Khatri et al., 2012).

2.1.2 Single-cell and spatial omics data

Recent advances in single-cell and spatial sequencing technologies have revolutionized cancer research by enabling high-resolution analyses of tumor biology (Zhang Y. et al., 2021; Lewis et al., 2021), revealing cellular heterogeneity and spatial organization within tumor microenvironment. While most of the related methods are still predominantly utilized in research settings, they offer a more detailed surrogate of the tumor profile, which cannot be achieved through bulk sequencing techniques. But these emerging technologies are increasingly poised to be applied in clinical settings.

Single-cell RNA sequencing (scRNA-seq) facilitates gene expression profiling at the individual cell level, uncovering rare cell populations, diverse cellular states, and dynamic transcriptional changes that bulk sequencing approaches might obscure (Macosko et al., 2015). This technique has been instrumental in characterizing tumor cell plasticity, immune infiltration, and resistance mechanisms across various cancers, such as the discovery of intra-tumoral heterogeneity in triple-negative breast cancer (Patel et al., 2014).

Spatial transcriptomics complements single-cell data by preserving the spatial context of gene expression. Platforms such as 10x Genomics Visium, Xenium, Nanostring GeoMx, CosMx, Slide-seq, and image mass cytometry enable the simultaneous measurement of gene expression and histological features, facilitating the mapping of transcriptional patterns to specific tissue regions (Ståhl et al., 2016; Nagasawa et al., 2024). Spatial transcriptomics has been instrumental in studying the tumor-immune microenvironment across various cancer types—including colorectal and gastric cancers, melanoma, breast cancer, non-small cell lung cancer, pancreatic cancer, and glioblastoma—where cellular interactions and tumor architecture are crucial for understanding disease progression. (Du et al., 2023; Nagasawa et al., 2024; Zhu et al., 2024; Boe et al., 2024; Lewis et al., 2021).

Moreover, combining traditional bulk omics data with single-cell and spatial omics can offer a more comprehensive view of tumor biology, supporting better-informed clinical decisions and personalized treatment strategies. By integrating these modalities, researchers can capture both the overall molecular landscape and the intricate details of cellular heterogeneity and spatial organization, uncovering nuances that are often masked in bulk analyses. This approach has been successfully demonstrated in studies combining TCGA data with single-cell approaches in melanoma (Tirosh et al., 2016) and ovarian cancer (Lähnemann et al., 2020).

In terms of AI application in single-cell and spatial transcriptomics data, several foundational models have been developed using extensive single-cell datasets, including scVI, scBERT, Geneformer, and scGPT (Lopez et al., 2018; Yang F. et al., 2022; Theodoris et al., 2023; Cui et al., 2024). These models have shown superior performance in tasks like cell type annotation and correcting batch effects (Kedzierska et al., 2025; Boiarsky et al., 2024). For spatial transcriptomics, generative models are broadly utilized to map spatial transcriptomics into latent space embeddings for more effective representation than the gene expression matrix (Chang et al., 2022; Long et al., 2023; Xu et al., 2024b). Clustering these derived embeddings helps identify distinct spatial domains from whole slide images. To identify spatially variable genes with expression patterns changing gradually across slides, Gaussian process models and Markov random fields are employed (Svensson et al., 2018; Sun et al., 2020; BinTayyash et al., 2025; Zhang et al., 2022). Additionally, graph neural networks are used to model cellular neighborhoods *via* graph structures, enabling the quantification of cell-cell interactions between different cell types (Yuan and Bar-Joseph, 2020; Fischer et al., 2023).

2.2 Images

2.2.1 Histology images

Histology images of stained tissue sections are central to cancer diagnosis: biopsy samples are fixed and stained (e.g., hematoxylin and eosin [H&E]) and examined by pathologists to assess malignancy, subtype, grade, and other histopathological features. Despite the routine digitization of whole-slide images (WSIs), these gigapixel datasets remain underutilized relative to genomic and other omics resources (Saltz et al., 2018). This landscape is rapidly changing - advances in deep learning (DL) models carry the promise to extract prognostic and predictive signals with clinical relevance directly from H&E slides (Saltz et al., 2018; Dang et al., 2025; Shamai et al., 2022). Indeed, several models have been proposed to aid pathologist to locate potential tumor tissue from H&E slides. Some of them (e.g., Paige Prostate (Rienda et al., 2024; Introducing FDA-Approved Paige Prostate, 2025) and Roche Digital Pathology Dx ("Roche Receives FDA Clearance on Its Digital Pathology Solution for Diagnostic Use, 2025)) also received FDA Breakthrough Device Designation. However, whole-slide images present unique bioinformatic challenges: a single WSI may yield hundreds of thousands of tiles, each exhibiting staining variability and often lacking region-level annotations (Gadermayr and Tschuchnig, 2024). To address these issues, contemporary pipelines first learn tile-level feature embeddings—using either supervised or self-supervised approaches—and then aggregate these embeddings into slide-level predictions *via* pooling or attention-based mechanisms (Xu et al., 2024a), as detailed in the following sections.

2.2.1.1 Fully-supervised learning

Fully supervised learning has pioneered the analysis of medical images in oncology, enabling tasks such as tumor detection and classification by training models on meticulously labeled datasets (Otálora et al., 2021; Turkki et al., 2015). In general, this class of method excels in utilizing annotated data to achieve good specificity and sensitivity. However, when dealing with commonly data models in histopathology, such as H&E stained images, fully supervised learning presents significant challenges (Lu et al., 2022; Rodriguez et al., 2022). Each input tile extracted from the source WSI requires explicit labeling, a process that is both labor-intensive and at times impractical. This requirement for fine-grained annotation restricts the availability and diversity of trainable datasets, ultimately limiting model generalizability across varied pathological contexts. Such limitations emphasize the emergence of self-supervised learning (SSL) methods as a superior alternative, offering the ability to

harness vast amounts of unlabeled data, to reduce dependency on exhaustive manual annotation, and to enhance the robustness and adaptability of models in digital pathology.

2.2.1.2 Self-supervised learning methods

Recent advances in SSL have introduced contrastive methods that effectively distinguish between positive and negative data pairs. These methods generate multiple augmented views of an image and use encoding techniques to enhance the similarity between related pairs while differentiating them from unrelated ones (Gui et al., 2024; Kang et al., 2023). Prominent examples such as SimCLR (Chen T. et al., 2020) and MoCo (Chen and He, 2021; He K. et al., 2020) achieve performance similar to fully-supervised methods without relying on labeled data, and have been successfully applied to pathological image processing (Ciga et al., 2022; Gadermayr and Tschuchnig, 2024). These methods are praised for their conceptual simplicity and modular design, often employing unique data augmentation techniques and a multilayer perceptron (MLP) as a projection head. However, a notable drawback is high computational demands. For SimCLR, large batch sizes are needed to provide diverse negative pairs, leading to high memory costs. To address these, implementations often approximate the loss by reducing the number of comparisons to random subsets during training (Chen et al., 2020). Another mitigation design can be found in MoCo, which features a mechanism for maintaining a dynamic dictionary of negative samples (He K. et al., 2020). This approach allows for the efficient use of extensive dictionaries without the necessity for large batch sizes, a feature that has been further refined in MoCo v2 (Chen and He, 2021).

Subsequently, non-contrastive methods have emerged with a significant improvement in operational efficiency. Typical methods of this class achieve learning by utilizing two neural networks, often configured as Siamese networks (Bromley et al., 1993), in an asymmetrical (Chen and He, 2021; Grill et al., 2020) or symmetrical model architecture (Zbontar et al., 2021). Unlike contrastive SSL methods, non-contrastive approaches focus exclusively on aligning positive view pairs augmented from the same image, also earning the designation "self-distilled." (Gui et al., 2024). With increasingly light-weight designs, notable examples of this method family further close the performance gap with, sometimes even exceed, their supervised learning benchmarks (Zbontar et al., 2021; Grill et al., 2020). Common challenges include trivial solutions, i.e., model weights collapsing to a constant during training. Therefore, popular implementations of non-contrastive methods primarily differ in their ways to avoid such trivial solutions, as discussed below.

BYOL (Bootstrap Your Own Latent) is the first method of this kind (Grill et al., 2020). This method employs two networks termed online and target networks. The online network consists of an encoder, a projector, and a predictor; while the target network, sharing the architecture of the online network, uses parameters which are a moving average of the online network's parameters. This design facilitates the online network's learning through self-supervised means, avoiding collapsing solutions through its momentum encoder. On the other hand, SimSiam (Simple Siamese Network) utilizes a pair of identical networks that share weights but apply stop-gradients to prevent trivial solutions (Chen and He, 2021). The constraint of dependency on asymmetric network design to prevent collapse in BYOL and SimSiam was relaxed by a more recent method, Barlow Twins (Zbontar et al., 2021). Inspired from neuroscientist Barlow (Barlow, 2012), this method applies redundancy reduction principles through identical networks. The core idea of avoiding collapsing solutions is to compute the cross-correlation matrix between outputs of the networks fed with distorted versions of an input, aligning this matrix as closely as possible to an identity matrix (Zbontar et al., 2021). It was pointed out, however, the Barlow Twins method can be sensitive to certain data augmentations, a trait shared with SimCLR but not observed in BYOL.

Clustering-based SSL methods offer a distinct approach by leveraging clustering algorithms to group data into meaningful representations. In general, these methods assign different views of input data to clusters and train the model to differentiate between clusters instead of individual representations, as seen in contrastive methods (Gui et al., 2024). By clustering features from unlabeled data, clustering methods allow models to understand the data distribution, obviating the need of large negative samples. For instance, DeepCluster (Caron et al., 2018) iterates between clustering image descriptors and updating network weights using cluster assignments as pseudo-labels, refining its learning in an end-to-end manner. As one of the founding SSL methods, DeepCluster is flexible with clustering algorithms like K-means, and supports various architectures [e.g., AlexNet, VGG (Krizhevsky, Sutskever, and Hinton, 2012; Simonyan and Zisserman, 2015)] to enhance feature transfer performance. Nevertheless, issues like empty clusters still haunt in early clustering methods; and their struggles with large datasets due to requiring full dataset passes for updates further limit their real-world applications. SwAW, a successor to DeepCluster, offers improvements by reformulating clustering-based SSL as an online method (Caron et al., 2020). Together with its novel multi-crop data augmentation strategy, SwAV enables scalable learning with smaller memory and computational requirements. Notably, SwAV shares the model designs of contrastive SSL by having "swapped" predictions, where the model predicts the cluster assignment of one view from the representation of another view. Despite reducing computational demands, it has been pointed out that SwAV has to carefully incorporate symmetry-breaking mechanisms to prevent trivial solutions, especially in scenarios with a high ratio of clusters to batch size (Zbontar et al., 2021).

Self-distillation methods leverage the concept of knowledge distillation (Hinton et al., 2015) for SSL. Knowledge distillation involves two models: a teacher model and a student model. The teacher model is typically large, complex, and extensively pre-trained on large datasets, while the student model is smaller and simpler. In the training process, the student model is trained to learn the output distribution (soft labels) of the teacher model instead of categorical classes (hard labels). This allows the student model to learn the rich representations encoded by the teacher model, enhancing its performance. In terms of self-distillation, there are two typical methods: Distillation with No Labels (DINO) (Caron et al., 2021) and Image BERT Pre-Training with Online Tokenizer (iBOT) (Zhou et al., 2022). DINO generates multiple crops of each image as different views. The original image is processed through the teacher model to learn a global representation, while the crops are processed through the student model to learn local representations.

The objective is to minimize the distance between global and local representations since they are from the same original image. iBOT were developed based on masked image modeling. During training, the student model learns to reconstruct masked crops under supervision of teacher models on same but unmasked crops.

A significant advantage of SSL is that it does not require labeled data, allowing for the full utilization of available datasets. As a result, researchers began to develop large models using extensive datasets curated from both public and proprietary sources, the so-called foundation models (Alfasly et al., 2024). Foundation models leverage SSL pretraining on massive amounts of data, enabling them to learn richer representations from input data, and achieve superior and more robust performance compared to fully-supervised models. Besides image-only SSL methods, some foundation models, inspired by vision-language models (Bordes et al., 2024), also take image-text pairs consisting of pathological images and corresponding captions as input (Huang et al., 2023; Ikezogwo et al., 2025; Lu et al., 2024a).

2.2.1.3 Slide-level aggregation

Since each whole slide image may contain a large number of tiles, it is essential to aggregate tile-level embeddings together for slide- or patient-level predictions, the so-called multi-instance learning (MIL) (Gadermayr and Tschuchnig, 2024). MIL focuses on predicting the label of a set of instances without knowing the label of each individual instance. Naïve methods such as averaging tile-level embeddings or predictions into slide-level embeddings or predictions, have been employed (Coudray et al., 2018; Kather et al., 2019). However, this approach treats all tiles equally and overlooks the fact that different tiles may exhibit distinct morphological features, leading to varying contributions to the prediction task. To address this, prior research has incorporated modules to identify the most important tiles and utilized only these top tiles for predictions (Campanella et al., 2019; Campanella et al., 2018; Courtiol et al., 2019). To leverage features from all tiles, attention-based methods have been developed, where an attention module is added to assign an attention score to each tile based on its embedding, enabling a weighted summation of all the tile embeddings (Carmichael et al., 2022; Ilse et al., 2018; Li et al., 2021b; Schirris et al., 2022).

With the recent advancements in transformer architecture, the self-attention (Vaswani et al., 2023) mechanism has been introduced to enhance the original attention mechanism (Li et al., 2021a; Shao et al., 2021; Zhao et al., 2022). The advantage of self-attention is that it accounts for pairwise correlations between tile-level embeddings, assigning attention to each tile in the context of all tiles rather than based on a single tile embedding. Furthermore, self-attention incorporates the spatial relationship of tiles through position encoding methods, allowing position information of all tiles to be encoded and passed to the model (Li et al., 2021a; Shao et al., 2021; Zhao et al., 2022). Besides self-attention, graph structures have also been used to capture spatial proximity between tiles (Chen et al., 2021b; Li et al., 2018). In this approach, a whole slide image is represented as a graph, where nodes represent tiles and edges reflect direct proximity between two tiles. Graph neural networks can then be used to aggregate node-level (tile-level) embeddings into a whole graph embedding (slide-level). Further research has also explored the application of self-attention mechanisms within graph structures (Ding S. et al., 2023; Zheng et al., 2022).

In addition, cluster-based methods have been employed to aggregate tile-level embeddings into slide-level embeddings (Lu et al., 2021; Yao et al., 2020; Yao et al., 2019). These methods first assign all tiles from a slide into several morphology-related clusters through unsupervised methods to reduce dimensionality. Next, they extract cluster-level embeddings for each cluster and aggregate all cluster-level embeddings into a slide-level embedding. This aggregation can be completed through simple concatenation or attention-based summation.

While most MIL methods are weakly-supervised as previously described, some studies have explored self-supervised MIL approaches to obtain slide-level embeddings without any labels. To achieve this, researchers extended the Vision Transformer (ViT) (Dosovitskiy et al., 2021) architecture to WSIs. Specifically, ViT processes an image of 256 × 256 pixels by cropping it into non-overlapping 16 × 16 patches. Then each patch is tokenized, and self-attention is calculated between tokens to derive the embedding of the original image. However, applying ViT to WSIs substantially increases computational costs due to the gigapixel size of WSIs, which can result in enormous tiles and make self-attention calculation impossible. To reduce computational costs, the Hierarchical Image Pyramid Transformer (HIPT) has been proposed (Chen et al., 2022a). HIPT breaks down the patching process into hierarchical levels, where the model first learns embeddings from small tiles, then progressively learns from larger tiles composed of these small tiles, ultimately learns the embeddings for the entire WSI. Prov-GigaPath (Xu et al., 2024a) leveraged the dilated attention (Ding J. et al., 2023) to replace the vanilla attention, reducing the computational complexity from quadratic to linear and enabling the attention calculations for billions of tokens (tiles). Since data labeling can be often costly and challenging to procure, these slide-level self-supervised learning techniques offer promising avenues for future research.

2.2.2 Radiology images

Unlike histopathology images, which requires tissue extraction, radiology imaging is a non-invasive technique that enables lesion detection on a tissue-wide scale and aids in clinical decision-making. CT and MRI are the two most prevalent data modalities within radiology. In terms of data structure, radiology images differ from histopathology images in following ways. Typically, radiology images are three dimensional, composed of several stacked 2D slices scanned at various locations within the human body, while histopathology images are usually 2D. Radiology images are smaller, with thousands of pixels per edge, whereas histopathology images usually contain gigapixels. Despite these differences, radiology images share several similar processing techniques with histopathology images. Initially, research focused on extracting hand-crafted features (e.g., lesion size) for clinical applications (Sousa et al., 2010; Riccardi et al., 2011; Ye et al., 2009). With advancements in deep learning, widely applied image-based deep learning models such as convolutional neural networks (CNNs) (Katzman et al., 2018; Wang et al., 2022a) and vision transformers (Murphy et al., 2022; Wollek et al., 2023) have been used to learn representative features from radiology images. Similar to histopathology, multi-instance learning approaches have been employed to integrate different radiology slices instead of tiles from whole slide images (Shin et al., 2019; Zhang et al., 2020).

TABLE 1 Summary of preprocessing methods across different modalities.

Preprocessing methods	Modality	Special user case
PCA, Gene signature, Pathway analysis	Omics	Feature extraction from multi-omics data
Single cell foundation model	Omics	Learning a better representation from single cell data
Gaussian process model, Markov random field	Omics	Identify spatial domains from spatial transcriptomics
Graph neural network	Omics	Decipher cell-cell interaction
Fully supervised learning	Images	End-to-end prediciton model development
Self-supervised learning	Images	Learning a better representation for images
Multi-instance learning	Images	Integrate regional features for whole-slide image level prediction
Traditional machine learning	EHR	Integrate common clinical factors for prediction modeling
Large language model	EHR	Curate unstructure clinical notes

Several studies have developed foundation models for radiology, showing promising performance in clinically relevant tasks such as nodule classification and survival prediction (Pai et al., 2024; Paschali et al., 2025; Wu et al., 2023). In terms of clinical application, radiomics-based AI models account for the highest proportion (~70%) of FDA approved AI tools till 2023, with common use cases including image reconstruction, tissue segmentation, and abnormal tissue detection (Joshi et al., 2024; Luchini et al., 2022).

2.3 Electronic health records

Electronic health records (EHRs) contain patient clinical information in both structured and unstructured data formats. Structured data is typically organized in tabular formats that include features such as diagnosis codes, laboratory test results, and lines of therapy for each entry. In contrast, unstructured data is more complex and often consists of clinical notes. Traditional machine learning methods, such as regression-based and kernel-based methods, have been employed to analyze structured data by correlating it with clinical outcomes (Daemen et al., 2012; Jarvis et al., 2013). More recently, neural networks have been widely used in embedding structured data into compact vector spaces to enhance predictive capabilities for clinical outcomes (Keyl et al., 2025; Rasmy et al., 2018; Shickel et al., 2018). For unstructured data, natural language processing (NLP) tools have been extensively used either to extract important information from clinical notes, converting them into structured data (Gholipour et al., 2023; Kehl et al., 2020), or to directly embed entire clinical notes into highly compact vector spaces for downstream predictions (Lu et al., 2024a; Xiang et al., 2025). Several companies have already integrated NLP tools into their clinical workflows to streamline data curation and enhance efficiency (Swaminathan et al., 2020). Taking together, we summarized above discussed preprocessing methods across different modalities in Table 1.

3 Multimodal integration

Despite recent advancements in biomarker discovery and data processing techniques in each modality, patients exhibiting similar biomarkers can still have distinct prognoses and treatment responses (Kern, 2012). This deviation can potentially be attributed to the extensive heterogeneity inherent in cancer, where each modality captures only a fragment of the entire tumor profile, thereby hindering precise patient stratification. Integrating multimodal data can offer complementary insights across modalities, facilitating a more comprehensive evaluation of the tumor profile. Considering the complex and diverse data structures across different modalities, deep learning becomes an optimal approach due to its advantage in processing and aligning high-dimensional and complex data. Due to the innate complexity of different modalities and deep learning algorithms, most multimodal fusion strategies are still being explored in research settings. Based on the data fusion stages, multimodal fusion strategies can be divided into early, late and intermediate fusion strategies (Figures 1A,B).

3.1 Early fusion

Early fusion refers to integrating multimodal data into a unified feature matrix at an initial stage, followed by the development of a single model based on the integrated feature matrix for given prediction tasks. Typically, features from different data modalities are directly concatenated together into the integrated feature matrix. Due to the significant divergence between certain data modalities, such as images and gene expression matrices, pre-trained unimodal models are commonly employed to extract dimensionally aligned feature vectors for each independent modality (Stahlschmidt et al., 2022). Note that, although multiple models may be utilized during early fusion, they are fixed or slightly adjusted for feature extraction purpose without further extensive training. The main model for development during early fusion is the one utilizing the integrated feature matrix as input.

Since only one model is focused for extensive training during early fusion, this strategy tends to be simple to design and implement, as it alleviates the development of multiple individual models. However, it does not consider the crosstalk between different modalities and treat each modality uniformly, which may result in suboptimal performance when dealing with complex data modalities (Huang et al., 2020a). Moreover, since

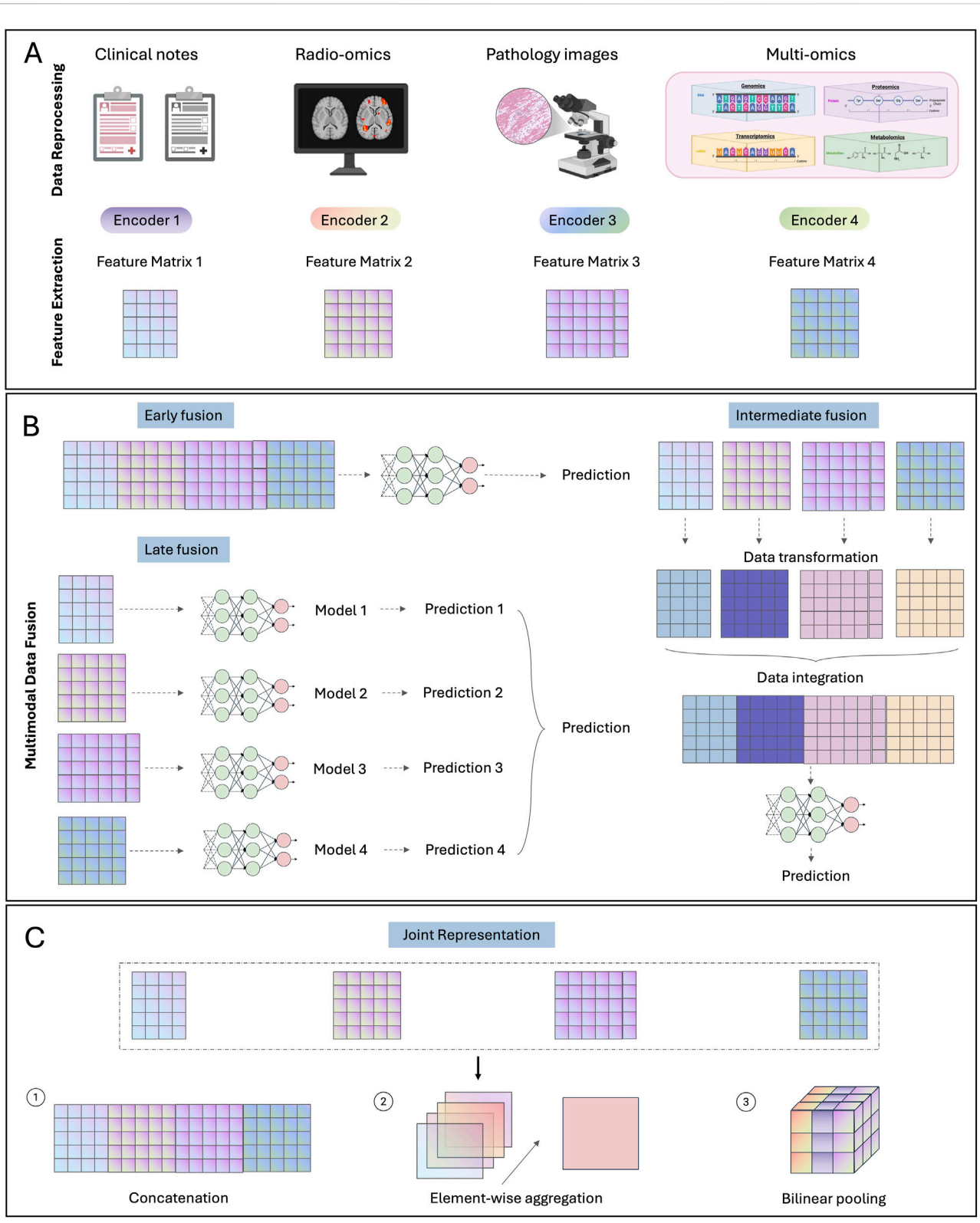

FIGURE 1
Multimodal fusion strategies. **(A)** Data preprocessing and feature extraction steps for common modalities. **(B)** Illustration of three multimodal fusion strategies: early, late and intermediate fusion. **(C)** Methods for generating joint representation.

the model is trained on the integrated feature, early fusion is not well-suited for addressing missing data.

Early fusion has been widely applied for cancer survival prediction. This includes integration of similar data structures such as different omics data (Bichindaritz et al., 2021; Zhao et al., 2021) and clinical table curated from EHR data (Xie et al., 2019). Further studies have expanded to incorporate more distinct modalities such as histology images (Chen et al., 2022c; Mobadersany et al., 2018) and spatial cellular patterns within these images (Chen et al., 2022b). Another key application area is cancer subtyping. Previous studies have integrated histology images with omics or clinical data for molecular subtyping in breast cancer (Yang J. et al., 2022). In above studies, pretrained CNN models were utilized to extract deep features from images for multimodal alignment. Besides deep features, some studies used hand craft features from image modality during early fusion. For example, Hyun et al. curated several quantitate features (e.g., skewness, kurtosis and texture features) from radiomic images, and then combined them with clinical features for histological subtyping in lung cancer (Hyun et al., 2019).

3.2 Late fusion

In contrast to early fusion occurring at the input level, late fusion focuses on the decision level. Specifically, late fusion develops customized models for each modality to obtain a high-level summarization for decision making, usually a numeric score for given prediction tasks. These scores from independent modalities are then aggregated into a single score for the final prediction.

Since each modality is represented by a score through its specifically designed model, late fusion can handle missing modalities by majority vote or averaging unimodal scores (Ramanathan et al., 2022). As these scores are calculated specifically for each modality, late fusion also allows customized processing of complex data structures to enhance model performance. Furthermore, the use of numerical scores simplifies multimodal alignment by reducing the need for processing all modal input into aligned dimensions for integration. However, developing customized models typically requires additional computing resources and specific domain knowledge for each modality, which may increase the complexity for model implementation and deployment. In addition, similar to early fusion, late fusion does not account for potential crosstalk between different data modalities (Huang et al., 2020a).

Conventionally, late fusion integrates scores from each modality. Previous studies have calculated scores for image and multi-omics data respectively, and combined the scores with clinical features for biomedical applications such as diagnosis (Reda et al., 2018), recurrence prediction (Rabinovici-Cohen et al., 2022), and prognosis estimation (Sun et al., 2019; Wang H. et al., 2021). Besides score-based approaches, some studies employed feature selection strategy for multi-omics data to identify key gene expression levels or copy number values instead of relying solely on one score (Arya and Saha, 2021). For image modality, hand crafted features, including nuclei area and shape, have been commonly used (Shao et al., 2020). With the recent advancement in deep learning, deep features extracted from customized deep learning models have increasingly been adopted for the image modality (Liu et al., 2022; Vale-Silva and Rohr, 2021).

3.3 Intermediate fusion

Intermediate fusion represents a bridging stage between early and late fusion. It does not directly combine input data like early fusion nor customize specific models for each data modality as late fusion. Instead, it focuses on incorporating interaction between different modalities, which has not been addressed by early and late-stage fusion, to align multimodal data and generate improved low-level multimodal feature representation. This method backpropagates the loss to the input from each modality, thereby enabling a dynamic integration of multimodal signals tailored to the specific prediction task (Huang et al., 2020a). This design allows intermediate fusion to effectively model complex interactions across multiple modalities. However, due to its commonly hyper-parameterized nature, intermediate fusion is prone to overfitting the training data, thereby limiting its generalizability (Huang et al., 2020b).

To model interactions across modalities, similarity- and attention-based methods are most commonly used. Similarity-based methods assume different modal data from the same patients to be closer to each other in latent space compared to those from different patients. Therefore, the objective function is typically designed to maximize the similarity between different modalities from the same patient while minimizing the similarity for those from different patients. In practice, most studies adopt cosine similarity due to its scale-invariance and simplicity (Cheerla and Gevaert, 2019; Radford et al., 2021). There are also some studies calculating the mean square error (MSE) as the measure of similarity (Ding K. et al., 2023).

Attention-based method is inspired by the Visual Question Answering (VQA) that learns the association between words and objects from sentences and images respectively, the so-called attention (Fukui et al., 2016). The self-attention mechanism from the transformer structure has been widely used (Vaswani et al., 2023). It involves projecting the input into Q (query), K (key) and V (Value) vectors to learn attention between two modality embeddings. Then the attention-weighted embeddings from each modality can be concatenated together as the final multimodal representations. Previous studies have used the genomic data to query the key and the values from image data for calculating the co-attention between two modalities, enabling identification of image regions specifically related to given molecular aberrations (Chen et al., 2021c). Further studies refined this approach by incorporating the reverse direction: querying genomic data based on image data (Jaume et al., 2024; Zuo et al., 2022). Besides, several studies employed optimal transport (Xu and Chen, 2023) or hierarchical fusion (Li et al., 2022) to compute the attention (interaction) between modalities.

Concatenation is commonly used for fusing features from different modalities (Figure 1C) (Mobadersany et al., 2018). To include feature interactions, element-wise aggregation operators such as element-wise summation or multiplication (Hadamard product) have been employed. More sophistically, bilinear pooling (Kronecker product), which models pairwise interactions

TABLE 2 Summary of multimodal integration studies.

Study	Modalities	Fusion strategy	Cancer type	Sample size	Performance
Bichindaritz et al.	mRNA, DNA methylation	Early	BRCA	1097	c-index: 0.72
Zhao et al.	mutation, mRNA, methylation, CNV	Early	pan-cancer	3400	c-index: 0.69
Xie et al.	mutation, mRNA, protein, CNV, clinical	Early	pan-cancer	5748	c-index: 0.53–0.84
Chen et al.	WSI, mRNA, mutation, CNV	Early	pan-cancer	5401	c-index: 0.64
Mobadersany et al.	WSI, mutation, CNV	Early	LGG, GBM	769	c-index: 0.78
Chen et al.	WSI, mRNA, mutation, CNV	Early	LGG, GBM, KIRC	1186	c-index: 0.78
Yang et al.	WSI, clinical	Early	BRCA	123	AUC: 0.72
Hyun et al.	PET/CT, clinical	Early	lung	396	AUC: 0.86, ACC: 0.77, F1: 0.77. precision: 0.80, recall: 0.75
Reda et al.	MRI, clinical	Late	PRAD	18	ACC: 0.78
Rabinovici-Cohen et al.	MRI, clinical	Late	BRCA	1738	AUC: 0.75, specificity: 0.57, sensitivity: 0.90
Sun et al.	mRNA, CNV, clinical	Late	BRCA	1980	AUC: 0.85
Wang et al.	CT, mRNA	Late	lung	130	AUC: 0.73
Arya et al.	mRNA, CNV, clinical	Late	BRCA	1980	AUC: 0.95
Shao et al.	WSI, mRNA, CNV, methylation	Late	KIRC, KIRP, LUSC	787	c-index: 0.76, AUC: 0.78
Liu et al.	WSI, mRNA, CNV	Late	BRCA	1098	AUC: 0.94, ACC: 0.88
Vale-Silva et al.	WSI, clinical, mRNA, microRNA, methylation, CNV	Late	pan-cancer	11315	c-index: 0.79
Cheerla et al.	WSI, clinical, mRNA, microRNA, methylation	Intermediate	pan-cancer	11160	c-index: 0.78
Ding et al.	WSI, mRNA, CNV, methylation	Intermediate	COADREAD	571	c-index: 0.71
Chen et al.	WSI, mRNA, mutation, CNV	Intermediate	BLCA, BRCA, GBMLGG, LUAD, UCEC	3523	c-index: 0.65
Jaume et al.	WSI, mRNA	Intermediate	BLCA, BRCA, STAD, COADREAD, HNSC	2233	c-index: 0.63
Zuo et al.	WSI, mRNA	Intermediate	BRCA	427	c-index: 0.74, AUC: 0.75
Xu et al.	WSI, mRNA, mutation, CNV	Intermediate	BLCA, BRCA, GBMLGG, LUAD, UCEC	2831	c-index: 0.71
Li et al.	WSI, mRNA, CNV, clinical	Intermediate	BRCA	1015	c-index: 0.77, AUC: 0.81
Wang et al.	WSI, mRNA	Intermediate	BRCA	345	c-index: 0.72, AUC: 0.82
Qiu et al.	WSI, mRNA, mutation, CNV	Intermediate	BLCA, KIRC, KIRP, LUSC, LUAD, PAAD	2250	c-index: 0.68

by calculating the outer product of two feature vectors, have been developed (Chen et al., 2022b; Chen et al., 2022c; Wang Z. et al., 2021). However, this operator usually results in a high-dimensional feature matrix. To reduce the high computational cost, factorized bilinear pooling methods have been developed based on low-rank matrix projections (Kim et al., 2017; Li et al., 2022a; Qiu et al., 2023).

Table 2 summarized above discussed studies. In practice, no fusion stage or operator consistently outperforms others across all scenarios. The choice of strategy should be guided by the specific data and prediction tasks.

4 Clinical application

In recent years, the rapid advancement of artificial intelligence (AI) and the explosive growth of multi-modal data have demonstrated remarkable potential of machine learning (ML) and deep learning (DL) in early cancer detection and diagnosis, molecular biomarker discovery and patient clinical outcome prediction (Figure 2). The following sections review key studies and recent progress for these three major application areas.

10.3389/fphar.2025.1609079

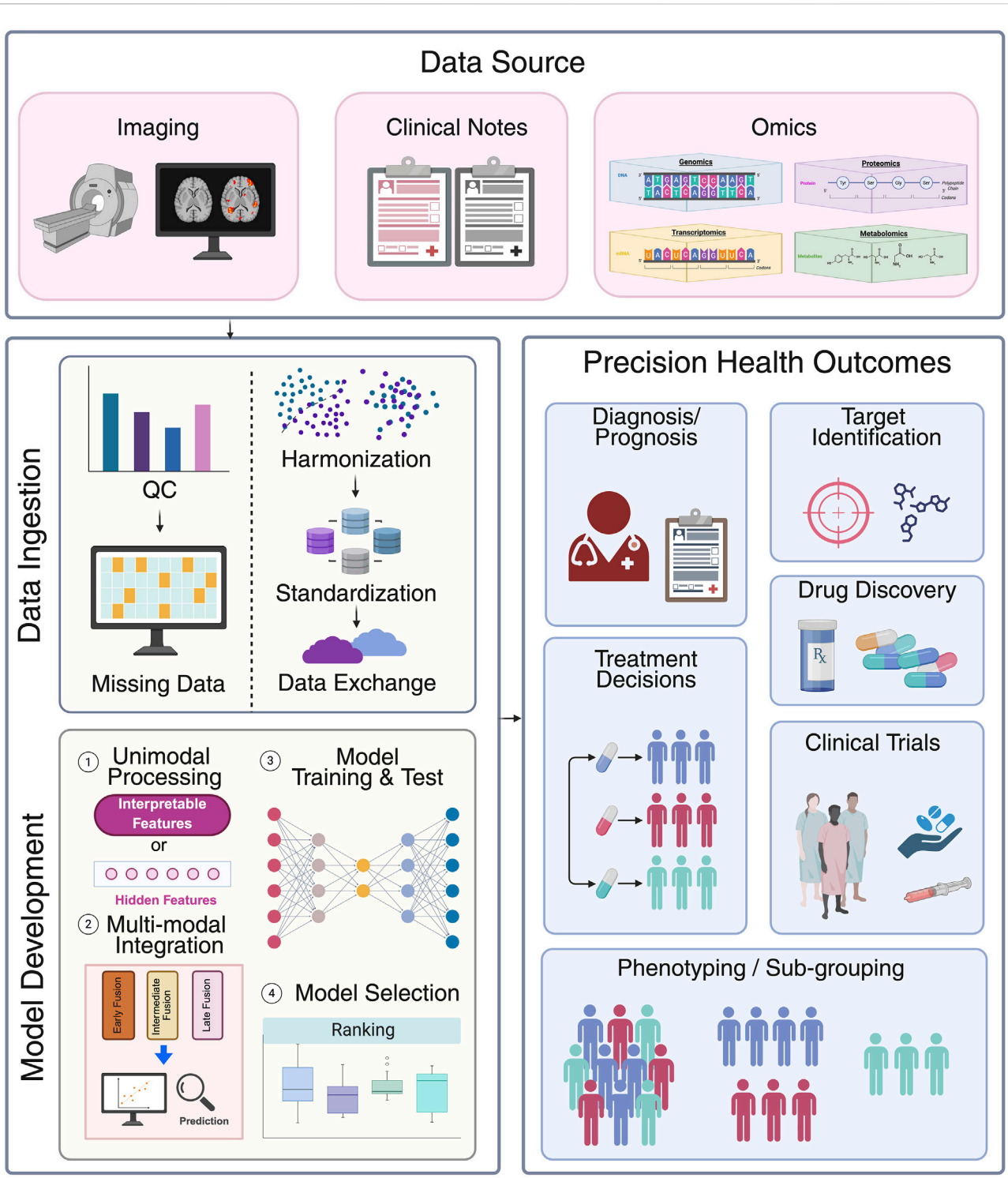

FIGURE 2
Data pipelines for multimodal data integration in clinical setting. First, various data modalities such as imaging, clinical notes and multi omics data are collected from data sources. Subsequentially, all data is ingested to data systems. This process involves quality control (e.g., presence of abnormal value or population prevalence), addressing missing data (e.g., discarding or imputing), harmonization (batch effects correction), standardization and loaded into systems for exchange. Model development involves four key steps: 1) unimodal processing: each data modality is first processed into interpretable features or hidden features via neural networks; 2) multi-modal integration: different data modalities are integrated together through early, intermediate, or late fusion strategy; 3) model training and test: different models are developed in curated datasets; and 4) model selection: The model with the best performance is selected to predict precision health outcomes such as diagnosis/prognosis, treatment decisions, patient sub-grouping, target identification, drug discovery, and to aid clinical trials. Created in BioRender. Wan, Z. (2025) https://BioRender.com/a7z643o.

4.1 Early detection/diagnosis

Early detection and diagnosis of cancers can significantly improve patient survival rates, treatment effectiveness, and patient life quality. Traditional cancer diagnosis typically involves non-invasive imaging (e.g., radiological scans), followed by invasive biopsy taken for histological examination if suspicious regions of tissues are detected (Díaz et al., 2024). However, these approaches usually only rely on a single-modality screening, which may miss early-stage tumors (Jiang et al., 2024; Park et al., 2022). They also suffer from false positives and delayed image evaluation by time-constrained physicians (Weiss et al., 2017). AI-driven models can leverage diverse data modalities to uncover hidden patterns and to increase sensitivity and accuracy to detect pre-malignant changes before symptoms appear for some cancers. Iniyan et al. combined various imaging modalities (e.g., mammograms, ultrasound, and MRI) and proposed a technique by employing a fusion joint transfer learning for breast cancer diagnosis. The proposed method demonstrated superior performance than single-modality approaches by using histopathological images and ultrasound (Iniyan et al., 2024). A novel AI-based predictive system (AIPS) was developed to integrate radiological imaging, clinical information and genomics data to improve personalized risk stratification and classification accuracy in lung cancer. The AIPS models can achieve AUCs ranging from 0.587 to 0.910 in detecting the location of lung nodules in a cost-effective manner by reducing resource-intensive steps such as manual image annotation and complex feature engineering (Batra et al., 2024).

Recent advancement in single-cell and spatial omics have revolutionized multi-modal data integration, enabling mapping of cell-specific gene expressions with the incorporation of spatial location information and associating with other traditional data modalities (Chen S. et al., 2024). Interestingly, Bae et al., developed a deep learning framework SPADE to identify important genes associated with morphological contexts by combining spatial transcriptomics data with overlapping high-resolution images. Both spatial gene expression and H&E imaging data were fed into a VGG-16 convolutional neural network (CNN) to extract imaging features based on spatial coordinates. The DL framework was applied for malignant vs. benign tissue classification and yielded an accuracy of 0.91, while the single-data modality with the accuracy dropping to 0.60 without imaging data (Bae et al., 2021).

The AI models with enhanced accuracy can alleviate the workload of pathologists for manual microscopic inspection and accelerating early detection of cancers, particularly in resource-limited settings.

4.2 Molecular biomarker discovery

Biomarkers play a crucial role in cancer risk stratification and personalized treatment design (Naik et al., 2020). The well-established molecular biomarkers include oncogenic mutations (e.g., TP53, BRAF V600 E, MYC amplification, etc.), therapeutic biomarkers (e.g., TMB, PD-L1, MSI status, tumor-infiltrating lymphocytes, etc.) as well as some emerging biomarkers (e.g., ctDNA, DNA methylation markers, etc.). The main advantage of multimodal data integration in biomarker discovery is to capture tumor complexity across different biological layers and to reveal shared associations across molecular alterations, tissue morphology and clinical attributes, aiming to provide a comprehensive molecular and phenotypic landscape of tumors (Lipkova et al., 2022; Llinas-Bertran et al., 2025; Yang et al., 2025).

One of the most successful multi-modal data integration applications is to fully leverage the easy accessibility of imaging data (e.g., H&E, CT scans, X-ray, MRI, etc.) in clinical setting on biomarker identification and prediction (Chiu and Yen, 2023; Mi et al., 2024; Pang et al., 2022). Unlike genomics sequencing or multi-omics profiling that requires specialized instruments and process as well as its high cost, imaging data is readily available and non-invasive in standardized cancer patient diagnosis and treatment clinical practice, making it a perfectly complementary resource to molecular data for biomarker discovery. The technical development of ML and DL further make imaging data as an efficiently and widely adopted tool by extracting quantitative morphological features to correlate with available patient molecular and clinical data.

One of the seminal studies by Coudray et al. (2018), demonstrated that H&E-stained WSIs can be used for lung cancer histological group classification and directly infer the most common mutations by a deep convolutional neural network (inception v3). The predictive models achieved promising performance with AUC from 0.73 to 0.85 for prediction of six lung cancer driver mutations. This study opened a new opportunity for integrating massive clinical imaging data into biomarker prediction, especially in the absence of omics data. More similar studies were followed for mutation prediction by imaging data for different cancer types (Bilal et al., 2021; Chen M. et al., 2020; Jang et al., 2020; Loeffler et al., 2022). More recently, some studies (Fu et al., 2020; Kather et al., 2020) attempted to predict any clinically actionable genetic alterations in pan-cancers by using different approaches, such as weakly supervised learning.

Besides mutations, AI technologies have also been applied to predict gene expression by using imaging data. Schmauch et al. (2020). Developed a novel deep learning algorithm HE2RNA by training TCGA datasets to predict mRNA expression directly from whole-slide histology images without the need for expert annotation. This method also provided visual spatialization of gene expression and it was validated by CD3- and CD20-stained samples. Moreover, this application can be expanded to spatial level. James Zou's lab used matched spatial transcriptomics data and H&E-stained histopathology images from breast cancer patients, enabling the prediction of spatial gene expression map directly from standard imaging slides (He B. et al., 2020). The developed DL algorithm ST-Net demonstrates robust generalization capability to other datasets, offering a cost-effective alternative to spatial transcriptomics.

Additional biomarkers such as TMB (Huang et al., 2020a; Jain and Massoud, 2020; Niu et al., 2022), MSI status (Gustav et al., 2024; Kather et al., 2019; Lee et al., 2021), PD-L1 (Li et al., 2024; Liang et al., 2024; Shamai et al., 2022), hormone-receptor status (Naik et al., 2020; Wu et al., 2024) have also been successfully predicted by imaging data leveraging AI technologies. These identified associated morphological features can be served as non-invasive biomarker surrogates when lacking extensive molecular profiling to guide patient tailored treatment strategy.

4.3 Prediction of patient prognosis and treatment response

Cancer patient prognosis and treatment response are usually only assessed by clinical variables; however, this trend is changing towards integrating multi-modal data to this end. The additional layers of information can provide a more comprehensive picture of underlying characteristics affecting patient survival and treatment response as well as the hidden relationships between these features. The estimate of patient clinical outcome has become crucial for physicians to monitor patient disease progression and to design effective therapeutic strategy.

Huang et al. (2019) introduced a deep learning framework called SALMON (survival analysis learning with multi-omics neural network) to incorporate diverse data types, including multi-omics and clinical information such as age and hormone receptor status. To solve the high dimensionality issue inherent in omics data, the authors first constructed co-expression networks to identify gene modules as eigengenes and investigate the contributions of these gene modules to the hazard ratio. This approach successfully reduced the feature space by approximately 99% and largely increased model robustness and effectiveness, leading to enhanced survival prognosis prediction for breast cancer patients. In addition, the gene modules identified as most significantly associated with the hazard ratio were further evaluated with pathway enrichment analysis to elucidate gene regulation mechanisms to enhance biological interpretation. Other similar studies using DL algorithms for patient prognosis prediction have been published in other cancer types (Boehm et al., 2022; Chaudhary et al., 2018; Steyaert et al., 2023).

The multi-modal integration can also be applied to predict treatment response by utilizing clinical trial data (Schweinar et al., 2024). This important study (Esteva et al., 2022) leveraged five phase III randomized trails, encompassing more than 5,000 prostate cancer patients with a median follow-up of 11.4 years. The team developed a multimodal deep learning model to predict long-term clinical outcome and achieved a 9.2%–14.6% relative improvement compared to risk stratification tools. Furthermore, this study integrated imaging data with clinical information, and provided pathologist interpretations on identified tissue clusters, demonstrating one of the main merits of multi-modal data integration than unimodal models–increased biological interpretation and clinically relevant inference.

The application has been expanded to the prediction of other clinical outcome such as recurrence (Lee et al., 2020; Yamamoto et al., 2019), drug side effects and toxicity (Men et al., 2019; Mukherjee et al., 2025) by incorporating diverse data modalities, all of which have significantly improved patient survival rates and optimized treatment strategy in precision oncology.

Moreover, the application is not only at the research setting, it has been also integrated into multidisciplinary tumor boards (MTBs) through the lens of clinicians—surgeons, medical oncologists, and radiation oncologists—emphasizing the clinically transformative potential (Nardone et al., 2024). Surgeons view AI as a tool for enhancing intraoperative decision-making and surgical education, with models like GoNoGoNet (Laplante et al., 2023) and DeepCVS (Mascagni et al., 2022) offering real-time anatomical guidance and safety assessments during procedures. Medical oncologists leverage AI for molecular profiling, treatment selection, and clinical trial optimization, with platforms such as Watson for Oncology and radiogenomic models predicting treatment responses and genetic mutations (Lee and Lee, 2020). Radiation oncologists benefit from AI in treatment planning and toxicity prediction, using tools like the Radiation Planning Assistant (RPA) to automate contouring and dose optimization (Court et al., 2018).

Interestingly, virtual technologies also play transformative roles in enhancing cancer diagnosis, patient treatment and support. For example, Zydowicz et al. introduced how 3D printing and augmented reality (AR)/virtual reality (VR) can improve surgical precision, reduce reoperation rates, and support rehabilitation and training (Żydowicz et al., 2024). Metaverse is an AI-integrated platform for immersive surgical planning and education, showing broader potential in healthcare (Żydowicz et al., 2024). These emerging AI tools require further clinical validation, ethical safeguards, and accessibility improvements to ensure safe and collaborative integration into existing clinical practices.

5 Challenges

5.1 Missing data

A major challenge in developing multimodal deep learning models for clinical application is dealing with missing data. For instance, molecular features might be missing for certain patients, as not all patients undergo genomic tests. Additionally, curating structured data from clinical notes is complicated, possibly leading to missing clinical variables. Survival data may also be missing or inaccurate when follow-up periods are limited. In cases where an entire modality (e.g., H&E slides) is often missing in certain patients, discarding the modality data for those patients can be an option (Jun-Ho and Lee, 2019). This approach is more compatible with late fusion since it combines scores from each modality, and the combination process is typically robust, regardless of the number of input scores. Conversely, if only a certain part of a modality (e.g., clinical variables) is missing, imputation methods can be used to estimate the missing values based on the available data (Luo Y. 2022; Yoon et al., 2019).

5.2 Multimodal alignment

Since different modalities can exhibit distinct dimensionalities, numerical scales, and data structures, it is necessary to align them into compatible formats before integration. With recent advances in vision-language models, the transformer architecture has become the paradigm for processing both image and language modalities into numerical representation vectors (Bordes et al., 2024). However, for molecular and structured clinical data with tabular formats, deep learning methods encounter significant challenges due to the highly different distributions across feature values (Grinsztajn et al., 2022). Traditional normalization methods are ineffective here, as gradient descent optimization and dropout regularization can disrupt this normalization, leading to instability in training, especially for deep neural networks (Grinsztajn et al., 2022). To

address this, self-normalizing neural networks (SNNs) have been introduced to preserve the data distribution at each layer (Klambauer et al., 2017). Further research has also incorporated biological knowledge into SNNs to enhance model performance (Jaume et al., 2024).

5.3 Insufficient interpretability

Despite impressive performance, deep learning models are challenging to interpret due to their hyper-parameterized structures. Previous studies have introduced several *post hoc* interpretation strategies for different modalities. For image modality, attention-mechanisms (Li et al., 2021a; Shao Z. et al., 2021) are used to identify tiles in whole slide images that significantly contribute to the model's output, with class activation maps (Afify et al., 2023; Selvaraju et al., 2017) employed to highlight important regions within each tile. For molecular data, methods based on Shapley Additive Explanation (SHAP) (Singh et al., 2017; Chen et al., 2022c) are utilized to determine the importance on each gene or pathway level. However, these interpretations remain abstract and are insufficient for drawing precise biological insights in clinical settings. Furthermore, in a multimodal context, interpretation becomes more complicated due to the necessity of disentangling the contributions from different modalities.

5.4 Data interoperability

Different institutions often maintain their own IT infrastructures for storing patient data, resulting in varied syntactic structures among different sources, which complicates efforts to centralize data. Furthermore, different workflows for curating these datasets introduce significant semantic variability across sources. As clinical guidelines are keep evolving, several clinical practice such as definition of cancer stage and standard of care can vary over time, hindering analysis of retrospective data. To support data sharing and centralization, several initiatives, including The Cancer Genome Atlas (TCGA), the Genomic Data Commons (GDC), the Database of Genotypes and Phenotypes (dbGAP), American Association for Cancer Research project Genomics Evidence Neoplasia Information Exchange (AACR project GENIE), The Cancer Imaging Archive (TCIA), the European Genome-phenome Archive (EGA), the Genomics Pathology Imaging Collection (GPIC), the Clinical Proteomic Tumor Analysis Consortium (CPTAC) are working towards standardizing data sources into a uniform structure (Weinstein et al., 2013; Zhang Z. et al., 2021; Tryka et al., n.d.; The AACR Project GENIE Consortium et al., 2017; Clark et al., 2013; Lappalainen et al., 2015; Jennings et al., 2022; Li Y. et al., 2023).

Another challenge for data sharing is the diversity of institutional privacy policies. Researchers must often undergo a series of legal reviews to comply with each institution's data-sharing policies, which can strongly delay the sharing process and hampers model development. Federated learning offers an alternative solution. This approach enables model development through distributed and decentralized systems, allowing institutions to comply with privacy laws and regulations without sharing confidential data directly (Rieke et al., 2020).

5.5 Clinician acceptance

Currently, one of the key challenges in the application of AI on multi-modal data from clinical perspective is clinician acceptance. Because of the "black box" nature of many multi-modal AI models described above, the interpretability and transparency of AI models have become the main concerns from clinicians, especially when they cannot fully understand or validate against their own expertise and experience. Additionally, discrepancies between AI-driven insights and established clinical workflows can create friction, particularly if the system's outputs are not seamlessly integrated into existing decision-making processes. Despite the rapid advancement of AI technologies, there remains gaps between their capabilities and alignment with clinical reasoning. Building clinician trust requires not only technical robustness but also thoughtful model design that prioritizes explainability, usability, and clinical translatability.

5.6 Generalizability of AI insights

Another significant limitation in applying AI to multi-modal data lies in the generalizability of the obtained AI results across diverse clinical settings and patient populations. Multi-modal models often rely on data from specific cohorts or geographic locations, which may not capture the full spectrum of variability in clinical practice, demographics, or disease presentation. As a result, models trained on one dataset may perform poorly when deployed in different environments, leading to difficulties of generalization of the obtained prediction results and AI insights. This issue is further compounded by the complexity of aligning and harmonizing heterogeneous data types—such as imaging, genomics, and electronic health records—which may be collected using different protocols or standards. Rigorous external validation and thorough consideration of data provenance are essential to overcoming key barriers to the widespread adoption of AI in real-world clinical workflows.

6 Discussion and outlook

6.1 Longitudinal multimodal fusion

Cancer is a dynamic process that evolves over time, driven by an intricate interplay of genetic, environment, and phenotypic changes. Most current cancer research is cross-sectional studies, which capture only a single snapshot of cancer at a specific timepoint (Avancini et al., 2024; Harle et al., 2020). In contrast, longitudinal data provides a more comprehensive perspective with patient temporal information across different disease phases, enabling the tracking of cancer progression and monitoring of treatment response (Zhuang et al., 2025). Harnessing AI-driven approaches to model longitudinal, multi-modal data presents a promising opportunity to increase predictive performance in cancer diagnosis, prognosis and clinical outcome. Some pioneering studies have explored various deep learning methods—such as recurrent neural network (RNN), transformers, self-supervised learning (SSL), reinforcement learning (RL) —by incorporating time embedding to analyze longitudinal data. However, most of these studies have focused on single-modality data, such as medical imaging (Gao et al., 2024; Gao

et al., 2025; Shen et al., 2023) or molecular profiling (Ballard et al., 2024; Zhang et al., 2024; Wekesa and Kimwele, 2023), without fully incorporating the wealth of information available across multiple modalities. This leaves significant untapped opportunity to leverage the strength of different data modalities for a more holistic and temporally resolved characterization of tumors (Zhuang et al., 2025).

Despite these advances, a major gap remains between current AI models and an ideal longitudinal fusion system. Most existing approaches primarily focus on relatively controlled datasets but struggle to align and integrate heterogeneous data types collected at irregular intervals—common in clinical practice—while also lack robust algorithm to handle missing data effectively or capture complex temporal dependencies across modalities (de Mortanges et al., 2024; Loni et al., 2025). An ideal system would seamlessly model time-aware interactions between different data modalities, dynamically adapting to continuously evolving patient status and treatment responses over time. Bridging this gap will require new analytical frameworks that are both modality-agnostic and temporally flexible, capable of learning robust representations from incomplete and asynchronous data streams.

Besides the methodologies, progresses in this area is also constrained by limited available public longitudinal, multimodal cancer datasets. The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Atlas Analysis Consortium (CPTAC) provide longitudinal clinical follow-up and survival outcomes, although molecular profiling is primarily cross-sectional. The National Lung Screening Trial (NLST) (National Lung Screening Trial Research Team, 2011) includes serial CT imaging and clinical metadata over multiple timepoints. These emerging resources offer solid foundation and partial solution. Developing well-annotated, longitudinal, multimodal datasets will be critical to enable reproducible research and advancing next-generation AI in oncology.

6.2 Integrate single-cell and spatial omics data

Compared to bulk-level omics data, single-cell and spatial omics data enable molecular characterization of individual cells within a population, providing deeper insights into complex cellular system in their spatial context (Du et al., 2024; Lee et al., 2024). Single-cell and spatial omics data have been unprecedentedly accumulated due to the rapid advancement of state-of-the-art sequencing technologies over the past decades. With the surge of AI/ML techniques—coupled with different fusion strategies described in this review—there is growing confidence in the potential of applying advanced AI/ML algorithms on these high-resolution data (Halawani et al., 2023). Such approaches have the power to enhance predictive performance and reveal novel biological insights. However, some challenges still remain (Athaya et al., 2023), for example, a lack of standardized workflow for data selection, preprocessing, normalization, and harmonization poses a significant barrier (Fan et al., 2020). In addition, the design of DL architecture needs to consider the high-dimensionality, sparsity and noise inherent in these data types (Stegle et al., 2015). Furthermore, integrating omics datasets across bulk, single cell and spatial levels with other data modalities remain a crucial challenge that must be addressed in future work to promote comprehensive cancer research.

6.3 Foundation models in healthcare

Foundation models are large generalist models pre-trained on extensive datasets to learn representations from vast amounts of information (Bommasani et al., 2022). They can be easily adapted to specific tasks through fine-tuning, often outperforming specialist models developed *via* fully supervised manner. These models show significant promise across various healthcare data modalities, including clinical notes, imaging data, and sequencing data. For instance, large language models like ChatCAD, MedAgents, and Med-PaLM can summarize clinical notes into reports and serve as virtual healthcare consultants (Tang J. et al., 2025; Tang et al., 2024; Singhal et al., 2023). Vision foundation models have been adapted for clinical tasks such as tissue segmentation, lesion detection, and survival prediction (Wang et al., 2022b; Chen R. J. et al., 2024; Xu Hang et al., 2024). Regarding sequencing data, the DNA foundation model Evo has been proposed to estimate gene essentiality (Nguyen et al., 2024). Protein language models like AlphaFold2/3, ESM3, and ProGen have been developed for predicting protein structure and function (Jumper et al., 2021; Abramson et al., 2024; Hayes et al., 2024; Madani et al., 2023). Foundation models like scVI, scGPT, and Geneformer have been utilized for cell type annotation and correcting batch effects in single-cell datasets (Lopez et al., 2018; Cui et al., 2024; Theodoris et al., 2023). By integrating data from different modalities, multimodal foundation models enable cross-modality applications and enhance performance beyond single-modality limitations. Vision-language models such as PathChat, LLaVA-Med, Clinical-BERT, and XrayGPT can generate clinical reports and provide insights based on clinical images (Lu et al., 2024b; Li C. et al., 2023; Huang Shih-Cheng et al., 2020; Thawakar et al., 2025). Among them, PathChat has received FDA Breakthrough Device Designation (PathChat, 2025). Molecular language models like BioMedGPT can analyze sequencing datasets given natural language queries (Zhang et al., 2024).

Despite the rapid advancements and promising potential of foundation models, several challenges persist. Training these models typically requires enormous amounts of data, which is often scarce in clinical settings due to difficulties in data curation and related ethics and privacy regulations (Bommasani et al., 2022; Willemink et al., 2020). The large architecture of these models and the extensive data requirements also pose challenges to computing infrastructures (Ding N. et al., 2023; Touvron et al., 2023). Additionally, due to the lack of grounding and potential biases in training data, AI models may generate fatally inaccurate results, the so-called AI hallucinations, which can be particularly dangerous when incorrect information is provided to patients without sufficient medical knowledge (Maleki et al., 2024). Thus, fostering data standardization, reducing computational costs, and ensuring safety controls are critical for the future development of foundation models.

6.4 Multimodal fusion strategy selection

Three strategies for multimodal fusion are discussed above, namely, early, late, and intermediate fusion (Huang Shih-Cheng

et al., 2020). Each strategy possesses distinct strengths and weaknesses, and no single strategy is universally optimal for all scenarios. Early fusion merges input features directly, making it simple to implement; however, it cannot handle missing modalities in the input data (Baltrusaitis et al., 2019). Conversely, late fusion customizes separate models to generate predictions for each available modality, addressing the missing modalities by averaging predictions. Despite its advantages, late fusion requires substantial computational costs for developing and implementing independent models (Baltrusaitis et al., 2019). Both early and late fusion strategies overlook the interactions between different modalities. Intermediate fusion tackles this issue by incorporating architectures that model inter-modality interactions, thereby extracting orthogonal features from each modality for improved prediction performance (Huang et al., 2020a). Nonetheless, this approach introduces additional parameters, making the model prong to overfit. Therefore, when deciding the fusion strategy, various factors should be considered, such as prediction task, computational resources, sample size, and proportion of missing values.

Author contributions

BZ: Writing – original draft, Writing – review and editing. ZW: Visualization, Writing – original draft. YL: Writing – original draft. XZ: Project administration, Writing – review and editing. JS: Writing – review and editing, Project administration. WZ: Writing – review and editing, Project administration. SW: Project administration, Writing – review and editing, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The design, study conduct, and financial support for this research were provided by AbbVie. No honoraria or payments were made for authorship.

Acknowledgments

Figure 2 created by Biorender.

Conflict of interest

All authors were employed by AbbVie.

The authors declare that this study received funding from AbbVie. The funder participated in the interpretation of data, review, and approval of the publication.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. During the preparation of this work the authors used GPT-4o and Claude-3-haiku to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Publisher's note

References

Aaltonen, L. A., Abascal, F., Abeshouse, A., Aburatani, H., Adams, D. J., Agrawal, N., et al. (2020). Pan-cancer analysis of whole genomes. *Nature* 578 (7793), 82–93. doi:10.1038/s41586-020-1969-6

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630 (8016), 493–500. doi:10.1038/s41586-024-07487-w

Afify, H. M., Mohammed, K. K., and Hassanien, A. E. (2023). Novel prediction model on OSCC histopathological images via deep transfer learning combined with grad-CAM interpretation. *Biomed. Signal Process. Control* 83 (May), 104704. doi:10.1016/j.bspc.2023.104704

Alfasly, S., Nejat, P., Hemati, S., Khan, J., Lahr, I., Alsaafin, A., et al. (2024). Foundation models for histopathology—fanfare or flair. *Mayo Clin. Proc. Digit. Health* 2 (1), 165–174. doi:10.1016/j.mcpdig.2024.02.003

Alter, O., Brown, P. O., and Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci.* 97 (18), 10101–10106. doi:10.1073/pnas.97.18.10101

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., et al. (2017). Classification of breast cancer histology images using convolutional neural networks. *PLOS ONE* 12 (6), e0177544. doi:10.1371/journal.pone.0177544

Arvaniti, E., Fricker, K. S., Moret, M., Rupp, N., Hermanns, T., Fankhauser, C., et al. (2018). Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Sci. Rep.* 8 (1), 12054. doi:10.1038/s41598-018-30535-1

Arya, N., and Saha, S. (2021). Multi-modal advanced deep learning architectures for breast cancer survival prediction. *Knowledge-Based Syst.* 221 (June), 106965. doi:10.1016/j.knosys.2021.106965

Athaya, T., Chowdhury Ripan, R., Li, X., and Hu, H. (2023). Multimodal deep learning approaches for single-cell multi-omics data integration. *Briefings Bioinforma.* 24 (5), bbad313. doi:10.1093/bib/bbad313

Avancini, A., Giannarelli, D., Borsati, A., Carnio, S., Cantale, O., Nepote, A., et al. (2024). A cross-sectional study evaluating the exercise discussion with oncologist during cancer consultation: the CONNECT study. *ESMO Open* 9 (7), 103624. doi:10.1016/j.esmoop.2024.103624

Bae, S., Choi, H., and Dong, S. L. (2021). Discovery of molecular features underlying the morphological landscape by integrating spatial transcriptomic data with deep features of tissue images. *Nucleic Acids Res.* 49 (10), e55. doi:10.1093/nar/gkab095

Ballard, J. L., Wang, Z., Li, W., Shen, Li, and Qi, L. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min.* 17 (1), 38. doi:10.1186/s13040-024-00391-z

Baltrusaitis, T., Ahuja, C., and Morency, L.-P. (2019). Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Analysis Mach. Intell.* 41 (2), 423–443. doi:10.1109/TPAMI.2018.2798607

Barlow, H. B. (2012). "Possible principles underlying the transformations of sensory messages," in *Sensory communication*. Editor W. A. Rosenblith (The MIT Press), 216–234. doi:10.7551/mitpress/9780262518420.003.0013

Batra, U., Nathany, S., Kaushik Nath, S., Jose, J. T., Sharma, T., Preeti, P., et al. (2024). AI-based pipeline for early screening of lung cancer: integrating radiology, clinical, and genomics data. *Lancet Regional Health - Southeast Asia* 24 (May), 100352. doi:10.1016/j.lansea.2024.100352

Bichindaritz, I., Liu, G., and Bartlett, C. (2021). Integrative survival analysis of breast cancer with gene expression and DNA methylation data. *Bioinformatics* 37 (17), 2601–2608. doi:10.1093/bioinformatics/btab140

Bilal, M., Ahmed Raza, S. E., Azam, A., Graham, S., Ilyas, M., Cree, I. A., et al. (2021). Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet Digital Health* 3 (12), e763–e772. doi:10.1016/S2589-7500(21)00180-1

BinTayyash, N., Georgaka, S., John, S. T., Ahmed, S., Boukouvalas, A., Hensman, J., et al. (2025). "Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments."37, 3788, 3795. doi:10.1093/bioinformatics/btab486

Boe, R. H., Triandafillou, C. G., Lazcano, R., Wargo, J. A., and Raj, A. (2024). Spatial transcriptomics reveals influence of microenvironment on intrinsic fates in melanoma therapy resistance. *bioRxiv*, 2024.06.30.601416. doi:10.1101/2024.06.30.601416

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 (2), 114–126. doi:10.1038/s41568-021-00408-3

Boiarsky, R., Singh, N. M., Buendia, A., Amini, A. P., Getz, G., and Sontag, D. (2024). Deeper evaluation of a single-cell foundation model. *Nat. Mach. Intell.* 6 (12), 1443–1446. doi:10.1038/s42256-024-00949-w

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., et al. (2022). On the opportunities and risks of foundation models. *arXiv*. doi:10.48550/arXiv.2108.07258

Bordes, F., Yuanzhe Pang, R., Ajay, A., Li, A. C., Bardes, A., Petryk, S., et al. (2024). An introduction to vision-language modeling. *arXiv*. doi:10.48550/arXiv.2405.17247

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1993). "Signature verification using a 'siamese' time delay neural network," in *Advances in neural information processing systems*, 6. Denver, CO: 7th NIPS Conference. Available online at: https://papers.neurips.cc/paper_files/paper/1993/hash/288cc0ff022877bd3df94bc9360b9c5d-Abstract.html.

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Silva, V. W. K., Busam, K. J., et al. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* 25 (8), 1301–1309. doi:10.1038/s41591-019-0508-1

Campanella, G., Silva, V. W. K., and Fuchs, T. J. (2018). Terabyte-scale deep multiple instance learning for classification and localization in pathology. *arXiv*. doi:10.48550/arXiv.1805.06983

Carmichael, I., Song, A. H., Chen, R. J., Williamson, D. F. K., Chen, T. Y., Mahmood, F., et al. (2022). "Incorporating intratumoral heterogeneity into weakly-supervised deep learning models via variance pooling," in *Medical image computing and computer assisted intervention – miccai 2022* (Cham: Springer Nature Switzerland), 387–397. doi:10.1007/978-3-031-16434-7_38

Caron, M., Bojanowski, P., Joulin, A., and Douze, M. (2018). "Deep clustering for unsupervised learning of visual features," in *Computer vision – eccv 2018. Lecture notes in computer science*. Editors V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer International Publishing), 11218, 139–156. doi:10.1007/978-3-030-01264-9_9

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., and Joulin, A. (2020). "Unsupervised learning of visual features by contrasting cluster assignments,", 33. Curran Associates, Inc, 9912–9924. doi:10.5555/3495724.3496555

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., et al. (2021). "Emerging properties in self-supervised vision transformers." *In Proceedings of the IEEE/CVF international conference on computer vision*, 9650–9660. [held virtually due to COVID].

Carrillo-Perez, F., Carlos Morales, J., Castillo-Secilla, D., Gevaert, O., Rojas, I., and Herrera, L. J. (2022). Machine-learning-based late fusion on multi-omics and multi-scale data for non-small-cell lung cancer diagnosis. *J. Personalized Med.* 12 (4), 601. doi:10.3390/jpm12040601

Chang, Y., He, F., Wang, J., Chen, S., Li, J., Liu, J., et al. (2022). Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. *Comput. Struct. Biotechnol. J.* 20 (January), 4600–4617. doi:10.1016/j.csbj.2022.08.029

Chaudhary, K., Poirion, O. B., Lu, L., and Garmire, L. X. (2018). Deep learning–based multi-omics integration robustly predicts survival in liver cancer. *Clin. Cancer Res.* 24 (6), 1248–1259. doi:10.1158/1078-0432.CCR-17-0853

Cheerla, A., and Gevaert, O. (2019). Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* 35 (14), i446–i454. doi:10.1093/bioinformatics/btz342

Chen, M., Zhang, B., Topatana, W., Cao, J., Zhu, H., Juengpanich, S., et al. (2020). Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *Npj Precis. Oncol.* 4 (1), 14–17. doi:10.1038/s41698-020-0120-3

Chen, R. J., Chen, C., Li, Y., Chen, T. Y., Trister, A. D., Krishnan, R. G., et al. (2022a). "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," in *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (New Orleans, LA, USA: IEEE, 16123–16134. doi:10.1109/CVPR52688.2022.01567

Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K., and Mahmood, F. (2021a). Synthetic data in machine learning for medicine and healthcare. *Nat. Biomed. Eng.* 5 (6), 493–497. doi:10.1038/s41551-021-00751-8

Chen, R. J., Lu, M. Y., Shaban, M., Chen, C., Chen, T. Y., Williamson, D. F. K., et al. (2021b). "Whole slide images are 2D point clouds: context-aware survival prediction using patch-based graph convolutional networks," in *Medical image computing and computer assisted intervention – miccai 2021*. Editors M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, et al. (Cham: Springer International Publishing), 339–349. doi:10.1007/978-3-030-87237-3_33

Chen, R. J., Lu, M. Y., Wang, J., Williamson, D. F. K., Rodig, S. J., Lindeman, N. I., et al. (2022b). Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans. Med. Imaging* 41 (4), 757–770. doi:10.1109/TMI.2020.3021387

Chen, R. J., Lu, M. Y., Weng, W.-H., Chen, T. Y., Williamson, D.Fk., Manz, T., et al. (2021c). "Multimodal Co-attention transformer for survival prediction in gigapixel whole slide images," in *2021 IEEE/CVF international conference on computer vision (ICCV)* (Montreal, QC, Canada: IEEE), 3995–4005. doi:10.1109/ICCV48922.2021.00398

Chen, R. J., Lu, M. Y., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Noor, Z., et al. (2022c). Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* 40 (8), 865–878.e6. doi:10.1016/j.ccell.2022.07.004

Chen, R. J., Tong, D., Lu, M. Y., Williamson, D. F. K., Jaume, G., Song, A. H., et al. (2024). Towards a general-purpose foundation model for computational pathology. *Nat. Med.* 30 (3), 850–862. doi:10.1038/s41591-024-02857-3

Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., et al. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat. Biomed. Eng.* 7 (6), 719–742. doi:10.1038/s41551-023-01056-8

Chen, S., Zhu, B., Huang, S., Hickey, J. W., Lin, K. Z., Snyder, M., et al. (2024). Integration of spatial and single-cell data across modalities with weakly linked features. *Nat. Biotechnol.* 42 (7), 1096–1106. doi:10.1038/s41587-023-01935-0

Chen, T., Simon, K., Norouzi, M., and Hinton, G. (2020). "A simple framework for contrastive learning of visual representations," in *Proceedings of the 37th international conference on machine learning* (Vienna, Austria: 37th ICML conference), 1597–1607. Available online at: https://proceedings.mlr.press/v119/chen20j.html.

Chen, X., and He, K. (2021). "Exploring simple siamese representation learning," in *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (USA: Nashville, TN), 15745–15753. doi:10.1109/CVPR46437.2021.01549

Chen, Z., Chen, Y., Sun, Yu, Tang, L., Zhang, Li, Hu, Y., et al. (2024b). Predicting gastric cancer response to anti-HER2 therapy or anti-HER2 combined immunotherapy based on multi-modal data. *Signal Transduct. Target. Ther.* 9 (1), 222–12. doi:10.1038/s41392-024-01932-y

Chiu, F.-Y., and Yen, Y. (2023). Imaging biomarkers for clinical applications in neuro-oncology: current status and future perspectives. *Biomark. Res.* 11 (1), 35. doi:10.1186/s40364-023-00476-7

Ciga, O., Xu, T., and Martel, A. L. (2022). Self supervised contrastive learning for digital histopathology. *Mach. Learn. Appl.* 7 (March), 100198. doi:10.1016/j.mlwa.2021.100198

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Paul, K., et al. (2013). The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J. Digital Imaging* 26 (6), 1045–1057. doi:10.1007/s10278-013-9622-7

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24 (10), 1559–1567. doi:10.1038/s41591-018-0177-5

Court, L. E., Kelly, K., McCarroll, R., Zhang, L., Yang, J., Simonds, H., et al. (2018). Radiation planning assistant - a streamlined, fully automated radiotherapy treatment planning system. *J. Vis. Exp. (JoVE)* (134 (April), e57411. doi:10.3791/57411

Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., et al. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat. Med.* 25 (10), 1519–1525. doi:10.1038/s41591-019-0583-3

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 39 (Suppl. l_1), D691–D697. doi:10.1093/nar/gkq1018

Cui, C., Asad, Z., Dean, W. F., Smith, I. T., Madden, C., Bao, S., et al. (2022). "Multi-modal learning with missing data for cancer diagnosis using histopathological and genomic data,", 12033. SPIE, 120331D–78. doi:10.1117/12.2612318. *Med. Imaging 2022 Computer-Aided Diagn.*

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., et al. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* 21 (8), 1470–1480. doi:10.1038/s41592-024-02201-0

Daemen, A., Timmerman, D., Van den Bosch, T., Bottomley, C., Kirk, E., Van Holsbeke, C., et al. (2012). Improved modeling of clinical data with kernel methods. *Artif. Intell. Med.* 54 (2), 103–114. doi:10.1016/j.artmed.2011.11.001

Dang, C., Qi, Z., Xu, T., Gu, M., Chen, J., Wu, J., et al. (2025). Deep learning–powered whole slide image analysis in cancer pathology. *Lab. Investig.* 105 (7), 104186. doi:10.1016/j.labinv.2025.104186

de Mortanges, P., Luo, H., Shu, S. Z., Kamath, A., Suter, Y., Mohamed, S., et al. (2024). Orchestrating explainable artificial intelligence for multimodal and longitudinal data in medical imaging. *Npj Digit. Med.* 7 (1), 1–10. doi:10.1038/s41746-024-01190-w

Diagnostic Use (2025). Roche Receives FDA clearance on its digital pathology solution for diagnostic use. *Diagnostics*. Available online at: https://diagnostics.roche.com/us/en/news-listing/2024/roche-receives-fda-clearance-on-its-digital-pathology-solution-for-diagnostic-use.html (Accessed June 9, 2025).

Diao, J. A., Wang, J. K., Fung Chui, W., Mountain, V., Chowdary Gullapally, S., Srinivasan, R., et al. (2021). Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat. Commun.* 12 (1), 1613. doi:10.1038/s41467-021-21896-9

Díaz, O., Rodríguez-Ruíz, A., and Sechopoulos, I. (2024). Artificial intelligence for breast cancer detection: technology, challenges, and prospects. *Eur. J. Radiology* 175 (June), 111457. doi:10.1016/j.ejrad.2024.111457

Ding, J., Ma, S., Dong, Li, Zhang, X., Huang, S., Wang, W., et al. (2023). LongNet: scaling transformers to 1,000,000,000 tokens. *arXiv*. doi:10.48550/arXiv.2307.02486

Ding, K., Zhou, Mu, Metaxas, D. N., and Zhang, S. (2023a). "Pathology-and-Genomics multimodal transformer for survival outcome prediction," in *Medical image computing and computer assisted intervention – miccai 2023*. Editors H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, et al. (Cham: Springer Nature Switzerland), 622–631. doi:10.1007/978-3-031-43987-2_60

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., et al. (2023b). Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nat. Mach. Intell.* 5 (3), 220–235. doi:10.1038/s42256-023-00626-4

Ding, S., Li, J., Wang, J., Ying, S., and Shi, J. (2023c). Multi-scale efficient graph-transformer for whole slide image classification. *IEEE J. Biomed. Health Inf.* 27 (12), 5926–5936. doi:10.1109/JBHI.2023.3317067

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). An image is worth 16x16 words: transformers for image recognition at scale. *arXiv*. doi:10.48550/arXiv.2010.11929

Du, J., Yang, Y.-C., An, Z.-J., Zhang, M.-H., Fu, X.-H., Huang, Z.-F., et al. (2023). Advances in spatial transcriptomics and related data analysis strategies. *J. Transl. Med.* 21 (1), 330. doi:10.1186/s12967-023-04150-2

Du, Y., Ding, X., and Ye, Y. (2024). The spatial multi-omics revolution in cancer therapy: precision redefined. *Cell Rep. Med.* 5 (9), 101740. doi:10.1016/j.xcrm.2024.101740

El, N., Omar, S. M., Loeffler, C. M. L., Carrero, Z. I., van Treeck, M., Kolbinger, F. R., et al. (2024). Regression-based deep-learning predicts molecular biomarkers from pathology slides. *Nat. Commun.* 15 (1), 1253. doi:10.1038/s41467-024-45589-1

Esteva, A., Feng, J., Wal, D. van der, Huang, S.-C., Simko, J. P., DeVries, S., et al. (2022). Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *Npj Digit. Med.* 5 (1), 1–8. doi:10.1038/s41746-022-00613-w

Fan, J., Slowikowski, K., and Zhang, F. (2020). Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp. and Mol. Med.* 52 (9), 1452–1465. doi:10.1038/s12276-020-0422-0

Fischer, D. S., Schaar, A. C., and Theis, F. J. (2023). Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* 41 (3), 332–336. doi:10.1038/s41587-022-01467-z

Fu, Yu, Jung, A. W., Torne, R. V., Gonzalez, S., Vöhringer, H., Shmatko, A., et al. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* 1 (8), 800–810. doi:10.1038/s43018-020-0085-8

Fukui, A., Park, D. H., Yang, D., Rohrbach, A., Darrell, T., and Rohrbach, M. (2016). Multimodal compact bilinear pooling for visual question answering and visual grounding. *arXiv*. doi:10.48550/arXiv.1606.01847

Gadermayr, M., and Tschuchnig, M. (2024). Multiple instance learning for digital pathology: a review of the state-of-the-art, limitations and future potential. *Comput. Med. Imaging Graph.* 112 (March), 102337. doi:10.1016/j.compmedimag.2024.102337

Gao, Y., Tan, T., Wang, X., Beets-Tan, R., Zhang, T., Han, L., et al. (2025). Multi-modal longitudinal representation learning for predicting neoadjuvant therapy response in breast cancer treatment. *IEEE J. Biomed. Health Inf.*, 1–10. doi:10.1109/JBHI.2025.3540574

Gao, Y., Ventura-Diaz, S., Wang, X., He, M., Xu, Z., Weir, A., et al. (2024). An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer. *Nat. Commun.* 15 (1), 9613. doi:10.1038/s41467-024-53450-8

Gholipour, M., Khajouei, R., Amiri, P., Gohari, S. H., and Ahmadian, L. (2023). Extracting cancer concepts from clinical notes using Natural Language processing: a systematic review. *BMC Bioinforma.* 24 (1), 405. doi:10.1186/s12859-023-05480-0

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., et al. (2020). Bootstrap Your own latent: a new approach to self-supervised learning. *arXiv*. doi:10.48550/arXiv.2006.07733

Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? *arXiv*. doi:10.48550/arXiv.2207.08815

Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., et al. (2024). "A survey on self-supervised learning: Algorithms, applications, and future trends. NW Washington, DC: IEEE Transactions on Pattern Analysis and Machine Intelligence. 46 (12) (2024): 9052–9071.

Gustav, M., Gabriel Reitsam, N., Carrero, Z. I., Loeffler, C. M. L., van Treeck, M., Yuan, T., et al. (2024). Deep learning for dual detection of microsatellite instability and POLE mutations in colorectal cancer histopathology. *Npj Precis. Oncol.* 8 (1), 115–11. doi:10.1038/s41698-024-00592-z

Halawani, R., Buchert, M., and Chen, Y.-P. P. (2023). Deep learning exploration of single-cell and spatially resolved cancer transcriptomics to unravel tumour heterogeneity. *Comput. Biol. Med.* 164 (September), 107274. doi:10.1016/j.compbiomed.2023.107274

Harle, A., Molassiotis, A., Buffin, O., Burnham, J., Smith, J., Yorke, J., et al. (2020). A cross sectional study to determine the prevalence of cough and its impact in patients with lung cancer: a patient unmet need. *BMC Cancer* 20 (1), 9. doi:10.1186/s12885-019-6451-1

Hayes, T., Rao, R., Akin, H., Sofroniew, N. J., Oktay, D., Lin, Z., et al. (2024). Simulating 500 million years of evolution with a language model. *bioRxiv*. doi:10.1101/2024.07.01.600583

He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å., et al. (2020). Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* 4 (8), 827–834. doi:10.1038/s41551-020-0578-x

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. *arXiv*. doi:10.48550/arXiv.1911.05722

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv*. doi:10.48550/arXiv.1503.02531

Huang, K., Altosaar, J., and Ranganath, R. (2020a). ClinicalBERT: modeling clinical notes and predicting hospital readmission. *arXiv*. doi:10.48550/arXiv.1904.05342

Huang, K., Lin, B., Liu, J., Liu, Y., Li, J., Tian, G., et al. (2022). Predicting colorectal cancer tumor mutational burden from histopathological images and clinical information using multi-modal deep learning. *Bioinformatics* 38 (22), 5108–5115. doi:10.1093/bioinformatics/btac641

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M. P. (2020b). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit. Med.* 3, 136. doi:10.1038/s41746-020-00341-z

Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T. J., and Zou, J. (2023). A visual–language foundation model for pathology image analysis using medical twitter. *Nat. Med.* 29 (9), 2307–2316. doi:10.1038/s41591-023-02504-3

Huang, Z., Zhan, X., Xiang, S., Johnson, T. S., Helm, B., Yu, C. Y., et al. (2019). SALMON: survival analysis learning with multi-omics neural networks on breast cancer. *Front. Genet.* 10 (March), 166. doi:10.3389/fgene.2019.00166

Hyun, S. H., Sun Ahn, Mi, Koh, Y. W., and Lee, Su J. (2019). A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin. Nucl. Med.* 44 (12), 956–960. doi:10.1097/RLU.0000000000002810

Ikezogwo, W. O., Saygin Seyfioglu, M., Ghezloo, F., Dylan, S. C. G., Mohammed, F. S., Kumar Anand, P., et al. (2025). Quilt-1M: one million image-text pairs for histopathology. *arXiv*. doi:10.48550/arXiv.2306.11207

Ilse, M., Tomczak, J. M., and Welling, M. (2018). Attention-based deep multiple instance learning. *arXiv*. doi:10.48550/arXiv.1802.04712

Iniyan, S., Senthil Raja, M., Poonguzhali, R., Vikram, A., Venkata Naga Ramesh, J., Mohanty, S. N., et al. (2024). Enhanced breast cancer diagnosis through integration of computer vision with fusion based joint transfer learning using multi modality medical images. *Sci. Rep.* 14 (1), 28376. doi:10.1038/s41598-024-79363-6

Introducing FDA-Approved Paige Prostate (2025). Introducing FDA-approved Paige prostate. Available online at: https://info.paige.ai/prostate (Accessed June 9, 2025).

Jain, M. S., and Massoud, T. F. (2020). Predicting tumour mutational burden from histopathological images using multiscale deep learning. *Nat. Mach. Intell.* 2 (6), 356–362. doi:10.1038/s42256-020-0190-5

Jané, P., Xu, X., Vincent, T., Jané, E., Karim, G., Dumont, R. A., et al. (2023). The imageable genome. *Nat. Commun.* 14 (1), 7329. doi:10.1038/s41467-023-43123-3

Jang, H.-J., Lee, A., Kang, J., Song, H., and Lee, S. H. (2020). Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning. *World J. Gastroenterology* 26 (40), 6207–6223. doi:10.3748/wjg.v26.i40.6207

Jarvis, S. W., Kovacs, C., Badriyah, T., Briggs, J., Mohammed, M. A., Meredith, P., et al. (2013). Development and validation of a decision tree early warning score based on routine laboratory test results for the discrimination of hospital mortality in emergency medical admissions. *Resuscitation* 84 (11), 1494–1499. doi:10.1016/j.resuscitation.2013.05.018

Jaume, G., Vaidya, A., Chen, R. J., Williamson, D. F. K., Liang, P.Pu, and Mahmood, F. (2024). "Modeling dense multimodal interactions between biological pathways and histology for survival prediction," in *2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Seattle, WA, USA: IEEE), 11579–11590. doi:10.1109/CVPR52733.2024.01100

Jennings, C. N., Humphries, M. P., Wood, S., Jadhav, M., Chabra, R., Brown, C., et al. (2022). Bridging the gap with the UK genomics pathology imaging collection. *Nat. Med.* 28 (6), 1107–1108. doi:10.1038/s41591-022-01798-z

Jensen, P. B., Jensen, L. J., and Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* 13 (6), 395–405. doi:10.1038/nrg3208

Jiang, B., Bao, L., He, S., Chen, X., Jin, Z., and Ye, Y. (2024). Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis. *Breast Cancer Res.* 26 (1), 137. doi:10.1186/s13058-024-01895-6

Jin, D., Liang, S., Shmatko, A., Arnold, A., Horst, D., Grünewald, T. G. P., et al. (2024). Teacher-student collaborated multiple instance learning for pan-cancer PDL1 expression prediction from histopathology slides. *Nat. Commun.* 15 (1), 3063. doi:10.1038/s41467-024-46764-0

Joshi, G., Jain, A., Reddy Araveeti, S., Adhikari, S., Garg, H., and Bhandari, M. (2024). FDA-approved artificial intelligence and machine learning (AI/ML)-Enabled medical devices: an updated landscape. *Electronics* 13 (3), 498. doi:10.3390/electronics13030498

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2

Jun-Ho, C., and Lee, J.-S. (2019). EmbraceNet: a robust deep learning architecture for multimodal classification. *Inf. Fusion* 51 (November), 259–270. doi:10.1016/j.inffus.2019.02.010

Kanavati, F., Toyokawa, G., Momosaki, S., Takeoka, H., Okamoto, M., Yamazaki, K., et al. (2021). A deep learning model for the classification of indeterminate lung carcinoma in biopsy whole slide images. *Sci. Rep.* 11 (1), 8110. doi:10.1038/s41598-021-87644-7

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Kang, M., Song, H., Park, S., Yoo, D., and Pereira, S. (2023). "Benchmarking self-supervised learning on diverse pathology datasets," in *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Vancouver, BC, Canada: IEEE), 3344–3354. doi:10.1109/CVPR52729.2023.00326

Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Echle, A., Muti, H. S., et al. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* 1 (8), 789–799. doi:10.1038/s43018-020-0087-6

Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., et al. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* 25 (7), 1054–1056. doi:10.1038/s41591-019-0462-y

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* 18 (1), 24. doi:10.1186/s12874-018-0482-1

Kedzierska, K. Z., Crawford, L., Amini, A. P., and Lu, A. X. (2025). Zero-shot evaluation reveals limitations of single-cell foundation models. *Genome Biol.* 26 (1), 101. doi:10.1186/s13059-025-03574-x

Kehl, K. L., Xu, W., Lepisto, E., Elmarakeby, H., Hassett, M. J., Van Allen, E. M., et al. (2020). Natural Language processing to ascertain cancer outcomes from medical oncologist notes. *JCO Clin. Cancer Inf.* 4 (August), 680–690. doi:10.1200/CCI.20.00020

Kern, S. E. (2012). Why Your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res.* 72 (23), 6097–6101. doi:10.1158/0008-5472.CAN-12-3232

Keyl, J., Keyl, P., Montavon, G., Hosch, R., Brehmer, A., Mochmann, L., et al. (2025). Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nat. Cancer* 6 (2), 307–322. doi:10.1038/s43018-024-00891-1

Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLOS Comput. Biol.* 8 (2), e1002375. doi:10.1371/journal.pcbi.1002375

Kim, J.-H., On, K.-W., Lim, W., Kim, J., Ha, J.-W., and Zhang, B.-T. (2017). Hadamard product for low-rank bilinear pooling. *arXiv*. doi:10.48550/arXiv.1610.04325

Klambauer, G., Unterthiner, T., Mayr, A., and Hochreiter, S. (2017). "Self-normalizing neural networks". In *Advances in neural information processing systems*. Long Beach, CA: Annual Conference on Neural Information Processing Systems, 30. (Accessed December 4-9, 2017).

Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., et al. (2022). Multimodal machine learning in precision health: a scoping review. *Npj Digit. Med.* 5 (1), 171–14. doi:10.1038/s41746-022-00712-8

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22 (3), 568–576. doi:10.1101/gr.129684.111

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems 25 (Lake Tahoe, NV) (Accessed December 3-6, 2012).

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21 (1), 31. doi:10.1186/s13059-020-1926-6

Laplante, S., Namazi, B., Kiani, P., Hashimoto, D. A., Alseidi, A., Pasten, M., et al. (2023). Validation of an artificial intelligence platform for the guidance of safe laparoscopic cholecystectomy. *Surg. Endosc.* 37 (3), 2260–2268. doi:10.1007/s00464-022-09439-9

Lappalainen, I., Almeida-King, J., Kumanduri, V., Alexander, S., John, D. S., ur-Rehman, S., et al. (2015). The European genome-phenome archive of human data consented for biomedical research. *Nat. Genet.* 47 (7), 692–695. doi:10.1038/ng.3312

Lee, B., Chun, S. H., Hong, Ji H., Woo, S., Kim, S., Jeong, J. W., et al. (2020). DeepBTS: prediction of recurrence-free survival of non-small cell lung cancer using a time-binned deep neural network. *Sci. Rep.* 10 (1), 1952. doi:10.1038/s41598-020-58722-z

Lee, K., and Lee, S. H. (2020). Artificial intelligence-driven oncology clinical decision support system for multidisciplinary teams. *Sensors* 20 (17), 4693. doi:10.3390/s20174693

Lee, S., Kim, G., Lee, J. Y., Lee, A. C., and Kwon, S. (2024). Mapping cancer biology in space: applications and perspectives on spatial omics for oncology. *Mol. Cancer* 23 (1), 26. doi:10.1186/s12943-024-01941-z

Lee, S. H., Song, H., and Hyun-Jong, J. (2021). Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer. *Int. J. Cancer* 149 (3), 728–740. doi:10.1002/ijc.33599

Lewis, S. M., Asselin-Labat, M.-L., Nguyen, Q., Jean, B., Tan, X., Wimmer, V. C., et al. (2021). Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods* 18 (9), 997–1012. doi:10.1038/s41592-021-01203-6

Li, Bo, Su, J., Liu, K., and Hu, C. (2024). Deep learning radiomics model based on PET/CT predicts PD-L1 expression in non-small cell lung cancer. *Eur. J. Radiology Open* 12 (June), 100549. doi:10.1016/j.ejro.2024.100549

Li, B., Yin, Li, and Eliceiri, K. W. (2021a). "Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning," in *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (USA: Nashville, TN), 14313–14323. doi:10.1109/CVPR46437.2021.01409

Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., et al. (2023). LLaVA-med: training a large language-and-Vision assistant for biomedicine in one day. *arXiv*. doi:10.48550/arXiv.2306.00890

Li, H., Yang, F., Zhao, Yu, Xing, X., Zhang, J., Gao, M., et al. (2021b). "DT-MIL: deformable transformer for multi-instance learning on histopathological image," in *Medical image computing and computer assisted intervention – miccai 2021*. Editors M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, et al. (Cham: Springer International Publishing), 206–216. doi:10.1007/978-3-030-87237-3_20

Li, R., Wu, X., Li, Ao, and Wang, M. (2022a). HFBSurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction. *Bioinformatics* 38 (9), 2587–2594. doi:10.1093/bioinformatics/btac113

Li, R., Yao, J., Zhu, X., Li, Y., and Huang, J. (2018). "Graph CNN for survival analysis on whole slide pathological images," in *Medical image computing and computer assisted intervention – miccai 2018*. Editors A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Cham: Springer International Publishing), 174–182. doi:10.1007/978-3-030-00934-2_20

Li, Y., Dou, Y., Leprevost, F.Da V., Geffen, Y., Calinawan, A. P., Aguet, F., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. *Cancer Cell* 41 (8), 1397–1406. doi:10.1016/j.ccell.2023.06.009

Li, Y. C., Wang, L., Law, J. N., Murali, T. M., and Pandey, G. (2022b). Integrating multimodal data through interpretable heterogeneous ensembles. *Bioinforma. Adv.* 2 (1), vbac065. doi:10.1093/bioadv/vbac065

Liang, C., Zheng, M., Zou, H., Han, Yu, Zhan, Y., Xing, Yu, et al. (2024). Deep learning-based image analysis predicts PD-L1 status from 18F-fdg PET/CT images in non-small-cell lung cancer. *Front. Oncol.* 14 (September), 1402994. doi:10.3389/fonc.2024.1402994

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lipkova, J., Chen, R. J., Chen, B., Lu, M. Y., Barbieri, M., Shao, D., et al. (2022). Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* 40 (10), 1095–1110. doi:10.1016/j.ccell.2022.09.012

Liu, T., Huang, J., Liao, T., Pu, R., Liu, S., and Peng, Y. (2022). A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *IRBM* 43 (1), 62–74. doi:10.1016/j.irbm.2020.12.002

Llinas-Bertran, A., Butjosa-Espín, M., Barberi, V., and Seoane, J. A. (2025). Multimodal data integration in early-stage breast cancer. *Breast* 80 (April), 103892. doi:10.1016/j.breast.2025.103892

Lobato-Delgado, B., Priego-Torres, B., and Sanchez-Morillo, D. (2022). Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis. *Cancers* 14 (13), 3215. doi:10.3390/cancers14133215

Loeffler, C. M. L., Ortiz Bruechle, N., Jung, M., Seillier, L., Rose, M., Laleh, N. G., et al. (2022). Artificial intelligence–based detection of FGFR3 mutational status directly from routine histology in bladder cancer: a possible preselection for molecular testing? *Eur. Urol. Focus* 8 (2), 472–479. doi:10.1016/j.euf.2021.04.007

Long, Y., Ang, K. S., Li, M., Chong, K. L. K., Sethi, R., Zhong, C., et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* 14 (1), 1155. doi:10.1038/s41467-023-36796-3

Loni, M., Poursalim, F., Asadi, M., and Gharehbaghi, A. (2025). A review on generative AI models for synthetic medical text, time series, and longitudinal data. *Npj Digit. Med.* 8 (1), 281–10. doi:10.1038/s41746-024-01409-w

Lopez, R., Jeffrey, R., Cole, M. B., Jordan, M. I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat. Methods* 15 (12), 1053–1058. doi:10.1038/s41592-018-0229-2

Lu, M. Y., Chen, B., Williamson, D. F. K., Chen, R. J., Liang, I., Ding, T., et al. (2024a). A visual-language foundation model for computational pathology. *Nat. Med.* 30 (3), 863–874. doi:10.1038/s41591-024-02856-4

Lu, M. Y., Chen, B., Williamson, D. F. K., Chen, R. J., Zhao, M., Chow, A. K., et al. (2024b). A multimodal generative AI copilot for human pathology. *Nature* 634 (8033), 466–473. doi:10.1038/s41586-024-07618-3

Lu, M. Y., Chen, R. J., Kong, D., Lipkova, J., Singh, R., Williamson, D. F. K., et al. (2022). Federated learning for computational pathology on gigapixel whole slide images. *Med. Image Anal.* 76 (February), 102298. doi:10.1016/j.media.2021.102298

Lu, M. Y., Williamson, D. F. K., 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 (6), 555–570. doi:10.1038/s41551-020-00682-w

Luchini, C., Pea, A., and Scarpa, A. (2022). Artificial intelligence in oncology: current applications and future perspectives. *Br. J. Cancer* 126 (1), 4–9. doi:10.1038/s41416-021-01633-1

Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., et al. (2022). BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings Bioinforma.* 23 (6), bbac409. doi:10.1093/bib/bbac409

Luo, Y. (2022). Evaluating the state of the art in missing data imputation for clinical data. *Briefings Bioinforma.* 23 (1), bbab489. doi:10.1093/bib/bbab489

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161 (5), 1202–1214. doi:10.1016/j.cell.2015.05.002

Madani, A., Krause, B., Greene, E. R., Subramanian, S., Mohr, B. P., Holton, J. M., et al. (2023). Large Language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* 41 (8), 1099–1106. doi:10.1038/s41587-022-01618-2

Maleki, N., Padmanabhan, B., and Dutta, K. (2024). "AI hallucinations: a misnomer worth clarifying," in *2024 IEEE conference on artificial intelligence (CAI)*, 133–138. doi:10.1109/CAI59869.2024.00033

Mardis, E. R. (2019). The impact of next-generation sequencing on cancer genomics: from discovery to clinic. *Cold Spring Harb. Perspect. Med.* 9 (9), a036269. doi:10.1101/cshperspect.a036269

Mascagni, P., Vardazaryan, A., Alapatt, D., Urade, T., Emre, T., Fiorillo, C., et al. (2022). Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. *Ann. Surg.* 275 (5), 955–961. doi:10.1097/SLA.0000000000004351

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 (2), 440–447. doi:10.1016/j.ijrobp.2019.06.009

Mi, H., Sivagnanam, S., Ho, W. J., Zhang, S., Bergman, D., Deshpande, A., et al. (2024). Computational methods and biomarker discovery strategies for spatial proteomics: a review in immuno-oncology. *Briefings Bioinforma.* 25 (5), bbae421. doi:10.1093/bib/bbae421

Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings Bioinforma.* 19 (6), 1236–1246. doi:10.1093/bib/bbx044

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Vega, J. E. V., et al. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci.* 115 (13), E2970-E2979–79. doi:10.1073/pnas.1717139115

Mukherjee, S., Swanson, K., Parker, W., Shivnaraine, R. V., Leitz, J., Pang, P. D., et al. (2025). ADMET-AI enables interpretable predictions of drug-induced cardiotoxicity. *Circulation* 151 (3), 285–287. doi:10.1161/CIRCULATIONAHA.124.070413

Murphy, Z. R., Venkatesh, K., Sulam, J., and Yi, P. H. (2022). Visual transformers and convolutional neural networks for disease classification on radiographs: a comparison of performance, sample efficiency, and hidden stratification. *Radiol. Artif. Intell.* 4 (6), e220012. doi:10.1148/ryai.220012

Nagasawa, S., Zenkoh, J., Suzuki, Y., and Suzuki, A. (2024). Spatial omics technologies for understanding molecular status associated with cancer progression. *Cancer Sci.* 115 (10), 3208–3217. doi:10.1111/cas.16283

Nagpal, K., Davis, F., Tan, F., Liu, Y., Chen, P.-H. C., Steiner, D. F., et al. (2020). Development and validation of a deep learning algorithm for gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol.* 6 (9), 1372–1380. doi:10.1001/jamaoncol.2020.2485

Naik, N., Ali, M., Esteva, A., Keskar, N. S., Press, M. F., Ruderman, D., et al. (2020). Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nat. Commun.* 11 (1), 5727. doi:10.1038/s41467-020-19334-3

Nardone, V., Marmorino, F., Germani, M. M., Cichowska-Cwalińska, N., Menditti, V. S., Gallo, P., et al. (2024). The role of artificial intelligence on tumor boards: perspectives from surgeons, medical oncologists and radiation oncologists. *Curr. Oncol.* 31 (9), 4984–5007. doi:10.3390/curroncol31090369

National Lung Screening Trial Research Team, Aberle, D. R., Berg, C. D., Black, W. C., Church, T. R., Fagerstrom, R. M., et al. (2011). The national lung screening trial: overview and study design. *Radiology* 258 (1), 243–253. doi:10.1148/radiol.10091808

Nguyen, E., Poli, M., Durrant, M. G., Kang, B., Katrekar, D., Li, D. B., et al. (2024). Sequence modeling and design from molecular to genome scale with Evo. *Science* 386 (6723), eado9336. doi:10.1126/science.ado9336

Niu, Yi, Wang, L., Zhang, X., Han, Yu, Yang, C., Bai, H., et al. (2022). Predicting tumor mutational burden from lung adenocarcinoma histopathological images using deep learning. *Front. Oncol.* 12 (June), 927426. doi:10.3389/fonc.2022.927426

Otálora, S., Marini, N., Müller, H., and Atzori, M. (2021). Combining weakly and strongly supervised learning improves strong supervision in gleason pattern classification. *BMC Med. Imaging* 21 (1), 77. doi:10.1186/s12880-021-00609-0

Pai, S., Bontempi, D., Hadzic, I., Prudente, V., Sokač, M., Chaunzwa, T. L., et al. (2024). Foundation model for cancer imaging biomarkers. *Nat. Mach. Intell.* 6 (3), 354–367. doi:10.1038/s42256-024-00807-9

Pang, Y., Wang, H., and Li, He (2022). Medical imaging biomarker discovery and integration towards AI-based personalized radiotherapy. *Front. Oncol.* 11 (January), 764665. doi:10.3389/fonc.2021.764665

Park, Ga E., Kang, B. J., Kim, S. H., and Lee, J. (2022). Retrospective review of missed cancer detection and its mammography findings with artificial-intelligence-based, computer-aided diagnosis. *Diagnostics* 12 (2), 387. doi:10.3390/diagnostics12020387

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., et al. (2023). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 41 (26), 4192–4199. doi:10.1200/JCO.22.02511

Paschali, M., Chen, Z., Blankemeier, L., Varma, M., Youssef, A., Bluethgen, C., et al. (2025). Foundation models in radiology: what, how, why, and why not. *Radiology* 314 (2), e240597. doi:10.1148/radiol.240597

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Sci. June* 344, 1396–1401. doi:10.1126/science.1254257

PathChat (2025). PathChat Receives FDA Breakthrough Device designation | modella AI. Available online at: https://www.modella.ai/pathchat-fda-breakthrough-designation (Accessed June 9, 2025).

Perou, C. M., Sørlie, T., Eisen, M. B., Rijn, M. van de, Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406 (6797), 747–752. doi:10.1038/35021093

Qiu, L., Khormali, A., and Liu, K. (2023). Deep biological pathway informed pathology-genomic multimodal survival prediction. *arXiv*. doi:10.48550/arXiv.2301.02383

Rabinovici-Cohen, S., Fernández, X. M., Rejo, B. G., Hexter, E., Cubelos, O. H., Pajula, J., et al. (2022). Multimodal prediction of five-year breast cancer recurrence in women who receive neoadjuvant chemotherapy. *Cancers* 14 (16), 3848. doi:10.3390/cancers14163848

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021). Learning transferable visual models from Natural Language supervision. *arXiv*. doi:10.48550/arXiv.2103.00020

Ramanathan, T. T., Hossen, J., and Sayeed, S. (2022). *Naïve bayes based multiple parallel fuzzy reasoning method for medical diagnosis*, 17.

Rasmy, L., Wu, Y., Wang, N., Geng, X., Zheng, W. J., Wang, F., et al. (2018). A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *J. Biomed. Inf.* 84 (August), 11–16. doi:10.1016/j.jbi.2018.06.011

Reda, I., Khalil, A., Elmogy, M., El-Fetouh, A. A., Ahmed, S., El-Ghar, M. A., et al. (2018). Deep learning role in early diagnosis of prostate cancer. *Technol. Cancer Res. and Treat.* 17 (January), 1533034618775530. doi:10.1177/1533034618775530

Riccardi, A., Sergueev Petkov, T., Ferri, G., Masotti, M., and Campanini, R. (2011). Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and zernike MIP classification. *Med. Phys.* 38 (4), 1962–1971. doi:10.1118/1.3560427

Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., et al. (2020). The future of digital health with federated learning. *Npj Digit. Med.* 3 (1), 119–7. doi:10.1038/s41746-020-00323-1

Rienda, I., Vale, J., Pinto, J., Polónia, A., and Eloy, C. (2024). "Using artificial intelligence to prioritize pathology samples: report of a test drive,". December. doi:10.1007/s00428-024-03988-1*Virchows Arch.*

Rodriguez, J. P. M., Rodriguez, R., Silva, V. W. K., Campos Kitamura, F., Corradi, G. C. A., and Rieder, R. (2022). Artificial intelligence as a tool for diagnosis in digital pathology whole slide images: a systematic review. *J. Pathology Inf.* 13 (January), 100138. doi:10.1016/j.jpi.2022.100138

Saltz, J., Gupta, R., Hou, Le, Kurc, T., Singh, P., Nguyen, Vu, et al. (2018). Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep

learning on pathology images. *Cell Rep.* 23 (1), 181–193.e7. doi:10.1016/j.celrep.2018.03.086

Schirris, Y., Gavves, E., Nederlof, I., Horlings, H. M., and Teuwen, J. (2022). DeepSMILE: contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. *Med. Image Anal.* 79 (July), 102464. doi:10.1016/j.media.2022.102464

Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., et al. (2020). A deep learning model to predict RNA-seq expression of tumours from whole slide images. *Nat. Commun.* 11 (1), 3877. doi:10.1038/s41467-020-17678-4

Schweinar, A., Wagner, F., Klingner, C., Festag, S., Cord, S., and Brodoehl, S. (2024). Simplifying multimodal clinical research data management: introducing an integrated and user-friendly database concept. *Appl. Clin. Inf.* 15 (March), 234–249. doi:10.1055/a-2259-0008

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Devi, P., and Batra, D. (2017). "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *2017 IEEE international conference on computer vision (ICCV)*, 618–626. doi:10.1109/ICCV.2017.74

Shamai, G., Livne, A., Polónia, A., Sabo, E., Cretu, A., Bar-Sela, G., et al. (2022). Deep learning-based image analysis predicts PD-L1 status from H&E-Stained histopathology images in breast cancer. *Nat. Commun.* 13 (1), 6753. doi:10.1038/s41467-022-34275-9

Shao, W., Han, Z., Cheng, J., Cheng, L., Wang, T., Sun, L., et al. (2020). Integrative analysis of pathological images and multi-dimensional genomic data for early-stage cancer prognosis. *IEEE Trans. Med. Imaging* 39 (1), 99–110. doi:10.1109/TMI.2019.2920608

Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al. (2021). "Transmil: transformer based correlated multiple instance learning for whole slide image classification,", in *Advances in neural information processing systems*. 34, 2136–2147. [NeurIPS 2021 is a Virtual-only Conference]. doi:10.5555/3540261.3540425

Shen, Y., Park, J., Yeung, F., Goldberg, E., Heacock, L., Shamout, F., et al. (2023). Leveraging transformers to improve breast cancer classification and risk assessment with multi-modal and longitudinal data. *arXiv*. doi:10.48550/arXiv.2311.03217

Shickel, B., James Tighe, P., Bihorac, A., and Rashidi, P. (2018). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inf.* 22 (5), 1589–1604. doi:10.1109/JBHI.2017.2767063

Shin, S. Y., Lee, S., Dong Yun, Il, Kim, S.Mi, and Lee, K.Mu (2019). Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. *IEEE Trans. Med. Imaging* 38 (3), 762–774. doi:10.1109/TMI.2018.2872031

Siavoshi, A., Taghizadeh, M., Dookhe, E., and Piran, M. (2022). Gene expression profiles and pathway enrichment analysis to identification of differentially expressed gene and signaling pathways in epithelial ovarian cancer based on high-throughput RNA-seq data. *Genomics* 114 (1), 161–170. doi:10.1016/j.ygeno.2021.11.031

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv*. doi:10.48550/arXiv.1409.1556

Singh, R., Lanchantin, J., Sekhon, A., and Qi, Y. (2017). Attend and predict: understanding gene regulation by selective attention on chromatin. *Adv. Neural Inf. Process. Syst.* 30 (December), 6785–6795. doi:10.5555/3295222.3295423

Singhal, K., Azizi, S., Tu, T., Sara Mahdavi, S., Wei, J., Chung, H. W., et al. (2023). Large Language models encode clinical knowledge. *Nature* 620 (7972), 172–180. doi:10.1038/s41586-023-06291-2

Sorin, V., Glicksberg, B. S., Artsi, Y., Barash, Y., Konen, E., Nadkarni, G. N., et al. (2024). Utilizing large language models in breast cancer management: systematic review. *J. Cancer Res. Clin. Oncol.* 150 (3), 140. doi:10.1007/s00432-024-05678-6

Sousa, S., da, J. R. F., Silva, A. C., Cardoso de Paiva, A., and Nunes, R. A. (2010). Methodology for automatic detection of lung nodules in computerized tomography images. *Comput. Methods Programs Biomed.* 98 (1), 1–14. doi:10.1016/j.cmpb.2009.07.006

Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Fernández Navarro, J., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82. doi:10.1126/science.aaf2403

Stahlschmidt, S. R., Benjamin, U., and Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: a review. *Briefings Bioinforma.* 23 (2), bbab569. doi:10.1093/bib/bbab569

Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16 (3), 133–145. doi:10.1038/nrg3833

Steyaert, S., Qiu, Y. L., Zheng, Y., Mukherjee, P., Vogel, H., and Gevaert, O. (2023). Multimodal deep learning to predict prognosis in adult and pediatric Brain tumors. *Commun. Med.* 3 (1), 44–15. doi:10.1038/s43856-023-00276-y

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sun, D., Wang, M., and Li, Ao (2019). A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 16 (3), 841–850. doi:10.1109/TCBB.2018.2806438

Sun, S., Zhu, J., and Xiang, Z. (2020). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* 17 (2), 193–200. doi:10.1038/s41592-019-0701-7

Svensson, V., Teichmann, S. A., and Stegle, O. (2018). SpatialDE: identification of spatially variable genes. *Nat. Methods* 15 (5), 343–346. doi:10.1038/nmeth.4636

Swaminathan, K. K., Mendonca, E., Mukherjee, P., Thirumalai, K., Newsome, R., and Narayanan, B. (2020). Development of an algorithm using Natural Language processing to identify metastatic breast cancer patients from clinical notes. *J. Clin. Oncol.* 38 (15_Suppl. l), e14056. doi:10.1200/JCO.2020.38.15_suppl.e14056

Tang, J., Xiao, H., Xiang, Li, Wang, W., and Gong, Z. (2025). "ChatCAD: an MLLM-guided framework for zero-shot CAD drawing restoration," in *Icassp 2025 - 2025 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 1–5. doi:10.1109/ICASSP49660.2025.10890248

Tang, X., Zou, A., Zhang, Z., Li, Z., Zhao, Y., Zhang, X., et al. (2024). MedAgents: large Language Models as collaborators for zero-shot medical reasoning. *arXiv*, 599–621. doi:10.18653/v1/2024.findings-acl.33

Tang, Z., Chen, G., Chen, S., Yao, J., You, L., and Chen, C. Y.-C. (2024). Modal-nexus auto-encoder for multi-modality cellular data integration and imputation. *Nat. Commun.* 15 (1), 9021. doi:10.1038/s41467-024-53355-6

Thawakar, O., Shaker, A., Shaji Mullappilly, S., Cholakkal, H., Anwer, R. M., Khan, S., et al. (2025). XrayGPT: chest radiographs summarization using medical vision-language models. *arXiv*. doi:10.48550/arXiv.2306.07971

The AACR Project GENIE Consortium, André, F., Arnedos, M., Baras, A. S., Baselga, J., Bedard, P. L., et al. (2017). AACR project GENIE: powering precision medicine through an international Consortium. *Cancer Discov.* 7 (8), 818–831. doi:10.1158/2159-8290.CD-17-0151

Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., et al. (2023). Transfer learning enables predictions in network biology. *Nature* 618 (7965), 616–624. doi:10.1038/s41586-023-06139-9

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, I. I. M. H., Treacy, D., Trombetta, J. J., et al. (2016). "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq,", 352. April, 189–196. doi:10.1126/science.aad0501*Science*

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., et al. (2023). LLaMA: open and efficient foundation language models. *arXiv*. doi:10.48550/arXiv.2302.13971

Tryka, K. A., Luning, H., Sturcke, A., Jin, Y., Wang, Z. Y., Ziyabari, L., et al. n.d. "NCBI's database of Genotypes and phenotypes: dbGaP.". doi:10.1093/nar/gkt1211

Turkki, R., Linder, N., Holopainen, T., Wang, Y., Grote, A., Lundin, M., et al. (2015). Assessment of tumour viability in human lung cancer xenografts with texture-based image analysis. *J. Clin. Pathology* 68 (8), 614–621. doi:10.1136/jclinpath-2015-202888

Unger, M., and Kather, J. N. (2024). A systematic analysis of deep learning in genomics and histopathology for precision oncology. *BMC Med. Genomics* 17 (1), 48. doi:10.1186/s12920-024-01796-9

Vale-Silva, L. A., and Rohr, K. (2021). Long-term cancer survival prediction using multimodal deep learning. *Sci. Rep.* 11 (1), 13505. doi:10.1038/s41598-021-92799-4

Valle, Í. F. do, Giampieri, E., Simonetti, G., Padella, A., Manfrini, M., Ferrari, A., et al. (2016). Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. *BMC Bioinforma.* 17 (12), 341. doi:10.1186/s12859-016-1190-7

Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., et al. (2024). Adapted large language models can outperform medical experts in clinical text summarization. *Nat. Med.* 30 (4), 1134–1142. doi:10.1038/s41591-024-02855-5

Varet, H., Brillet-Guéguen, L., Coppée, J.-Y., and Dillies, M.-A. (2016). SARTools: a DESeq2-and EdgeR-based R pipeline for comprehensive differential analysis of RNA-seq data. *PLOS ONE* 11 (6), e0157022. doi:10.1371/journal.pone.0157022

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2023). Attention is all you need. *arXiv*. doi:10.48550/arXiv.1706.03762

Wan, T., Cao, J., Chen, J., and Qin, Z. (2017). Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. *Neurocomputing, Adv. Comput. Tech. big Med. image data* 229 (March), 34–44. doi:10.1016/j.neucom.2016.05.084

Wang, H., Subramanian, V., and Syeda-Mahmood, T. (2021). *Modeling uncertainty in multi-modal fusion for lung cancer survival analysis*. IEEE 18th International Symposium on Biomedical Imaging ISBI, 1169–1172. doi:10.1109/ISBI48211.2021.9433823

Wang, S., Yu, He, Gan, Y., Wu, Z., Li, E., Li, X., et al. (2022a). Mining whole-lung information by artificial intelligence for predicting egfr genotype and targeted therapy response in lung cancer: a multicohort study. *Lancet Digital Health* 4 (5), e309–e319. doi:10.1016/S2589-7500(22)00024-3

Wang, Xi, Jiang, Y., Chen, H., Zhang, T., Han, Z., Chen, C., et al. (2023). Cancer immunotherapy response prediction from multi-modal clinical and image data using semi-supervised deep learning. *Radiotherapy Oncol.* 186 (September), 109793. doi:10.1016/j.radonc.2023.109793

Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., et al. (2022b). Transformer-based unsupervised contrastive learning for histopathological image classification. *Med. Image Anal.* 81 (October), 102559. doi:10.1016/j.media.2022.102559

Wang, Z., Li, R., Wang, M., and Li, Ao (2021). GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction. *Bioinformatics* 37 (18), 2963–2970. doi:10.1093/bioinformatics/btab185

Waqas, A., Tripathi, A., Ramachandran, R. P., Stewart, P. A., and Rasool, G. (2024). Multimodal data integration for oncology in the era of deep neural networks: a review. *Front. Artif. Intell.* 7 (July), 1408843. doi:10.3389/frai.2024.1408843

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., et al. (2013). The cancer genome Atlas pan-cancer analysis project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764

Weiss, J. E., Goodrich, M., Harris, K. A., Chicoine, R. E., Synnestvedt, M. B., Pyle, S. J., et al. (2017). Challenges with identifying indication for examination in breast imaging as a key clinical attribute in practice, research, and policy. *J. Am. Coll. Radiology* 14 (2), 198–207.e2. doi:10.1016/j.jacr.2016.08.017

Wekesa, J. S., and Kimwele, M. (2023). A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. *Front. Genet.* 14 (July), 1199087. doi:10.3389/fgene.2023.1199087

Weronika Magdalena, Z., Jaroslaw, S., and Luigi, M. (2024b). Navigating the metaverse: a new virtual tool with promising real benefits for breast cancer patients. *J. Clin. Med.* 13 (15), 4337. doi:10.3390/jcm13154337

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., et al. (2020). Preparing medical imaging data for machine learning. *Radiology* 295 (1), 4–15. doi:10.1148/radiol.2020192224

Wollek, A., Graf, R., Čečatka, S., Fink, N., Willem, T., Sabel, B. O., et al. (2023). Attention-based saliency maps improve interpretability of pneumothorax classification. *Radiol. Artif. Intell.* 5 (2), e220187. doi:10.1148/ryai.220187

Wu, C., Zhang, X., Zhang, Ya, Wang, Y., and Xie, W. (2023). Towards generalist foundation model for radiology by leveraging web-scale 2D&3D medical data. doi:10.48550/arXiv.2308.02463

Wu, J., Ge, L., Guo, Y., Zhao, A., Yao, J., Wang, Z., et al. (2024). Predicting hormone receptor status in invasive breast cancer through radiomics analysis of long-Axis and short-Axis ultrasound planes. *Sci. Rep.* 14 (1), 16503. doi:10.1038/s41598-024-67145-z

Xiang, J., Wang, X., Zhang, X., Xi, Y., Eweje, F., Chen, Y., et al. (2025). A vision–language foundation model for precision oncology. *Nature* 638 (8051), 769–778. doi:10.1038/s41586-024-08378-w

Xie, G., Dong, C., Kong, Y., Zhong, J. F., Li, M., and Wang, K. (2019). Group lasso regularized deep learning for cancer prognosis from multi-omics and clinical features. *Genes* 10 (3), 240. doi:10.3390/genes10030240

Xu, H., Fu, H., Long, Y., Ang, K. S., Sethi, R., Chong, K., et al. (2024a). Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med.* 16 (1), 12. doi:10.1186/s13073-024-01283-x

Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., et al. (2024b). A whole-slide foundation model for digital pathology from real-world data. *Nature* 630 (8015), 181–188. doi:10.1038/s41586-024-07441-w

Xu, Y., and Chen, H. (2023). "Multimodal optimal transport-based Co-attention transformer with global structure consistency for survival prediction," in *2023 IEEE/CVF international conference on computer vision (ICCV)* (Paris, France: IEEE), 21184–21194. doi:10.1109/ICCV51070.2023.01942

Yamamoto, Y., Tsuzuki, T., Akatsuka, J., Ueki, M., Morikawa, H., Numata, Y., et al. (2019). Automated acquisition of explainable knowledge from unannotated histopathology images. *Nat. Commun.* 10 (1), 5642. doi:10.1038/s41467-019-13647-8

Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., et al. (2022). scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat. Mach. Intell.* 4 (10), 852–866. doi:10.1038/s42256-022-00534-z

Yang, H., Yang, M., Chen, J., Yao, G., Zou, Q., and Jia, L. (2025). Multimodal deep learning approaches for precision oncology: a comprehensive review. *Briefings Bioinforma.* 26 (1), bbae699. doi:10.1093/bib/bbae699

Yang, J., Ju, J., Guo, L., Ji, B., Shi, S., Yang, Z., et al. (2022). Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput. Struct. Biotechnol. J.* 20 (January), 333–342. doi:10.1016/j.csbj.2021.12.028

Yao, J., Zhu, X., and Huang, J. (2019). "Deep multi-instance learning for survival prediction from whole slide images," in *Medical image computing and computer assisted intervention – miccai 2019*. Editors D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, et al. (Cham: Springer International Publishing), 496–504. doi:10.1007/978-3-030-32239-7_55

Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., and Huang, J. (2020). Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med. Image Anal.* 65 (October), 101789. doi:10.1016/j.media.2020.101789

Ye, X., Lin, X., Dehmeshki, J., Slabaugh, G., and Beddoe, G. (2009). Shape-based computer-aided detection of lung nodules in thoracic CT images. *IEEE Trans. Biomed. Eng.* 56 (7), 1810–1820. doi:10.1109/TBME.2009.2017027

Yoon, J., Zame, W. R., and Schaar, M. van der (2019). Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Trans. Bio-Medical Eng.* 66 (5), 1477–1490. doi:10.1109/TBME.2018.2874712

Yuan, Ye, and Bar-Joseph, Z. (2020). GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol.* 21 (1), 300. doi:10.1186/s13059-020-02214-w

Zbontar, J., Jing, Li, Misra, I., LeCun, Y., and Deny, S. (2021). "Barlow twins: Self-supervised learning via redundancy reduction," in *International conference on machine learning*. PMLR.

Zhang, B., Li, C., Wu, J., Zhang, J., and Cheng, C. (2024). DeepCG: a cell graph model for predicting prognosis in lung adenocarcinoma. *Int. J. Cancer* 154 (12), 2151–2161. doi:10.1002/ijc.34901

Zhang, Ke, Feng, W., and Wang, P. (2022). Identification of spatially variable genes with graph cuts. *Nat. Commun.* 13 (1), 5488. doi:10.1038/s41467-022-33182-3

Zhang, K., Zhou, R., Adhikarla, E., Yan, Z., Liu, Y., Yu, J., et al. (2024a). BiomedGPT: a generalist vision-language foundation model for diverse biomedical tasks. *Nat. Med.* 30 (11), 3129–3141. doi:10.1038/s41591-024-03185-2

Zhang, S., Yuan, J., Sun, Yu, Wu, F., Liu, Z., Zhai, F., et al. (2024b). Machine learning on longitudinal multi-modal data enables the understanding and prognosis of alzheimer's disease progression. *iScience* 27 (7), 110263. doi:10.1016/j.isci.2024.110263

Zhang, Xi, Lu, Di, Gao, P., Tian, Q., Lu, H., Xu, X., et al. (2020). Survival-relevant high-risk subregion identification for glioblastoma patients: the MRI-based multiple instance learning approach. *Eur. Radiol.* 30 (10), 5602–5610. doi:10.1007/s00330-020-06912-8

Zhang, Y., Peng, C., Wang, Q., Song, D., Li, K., and Kevin Zhou, S. (2025). Unified multi-modal image synthesis for missing modality imputation. *IEEE Trans. Med. Imaging* 44 (1), 4–18. doi:10.1109/TMI.2024.3424785

Zhang, Y., Wang, D., Peng, M., Tang, Le, Ouyang, J., Xiong, F., et al. (2021a). Single-cell RNA sequencing in cancer research. *J. Exp. and Clin. Cancer Res.* 40 (1), 81. doi:10.1186/s13046-021-01874-1

Zhang, Z., Hernandez, K., Savage, J., Li, S., Miller, D., Agrawal, S., et al. (2021). Uniform genomic data analysis in the NCI genomic data commons. *Nat. Commun.* 12 (1), 1226. doi:10.1038/s41467-021-21254-9

Zhao, C., Su, K.-J., Wu, C., Cao, X., Sha, Q., Wu, Li, et al. (2024). Multi-scale variational autoencoder for imputation of missing values in untargeted metabolomics using whole-genome sequencing data. *Comput. Biol. Med.* 179 (September), 108813. doi:10.1016/j.compbiomed.2024.108813

Zhao, L., Dong, Q., Luo, C., Wu, Y., Bu, D., Qi, X., et al. (2021). DeepOmix: a scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis. *Comput. Struct. Biotechnol. J.* 19 (January), 2719–2725. doi:10.1016/j.csbj.2021.04.067

Zhao, Yu, Lin, Z., Sun, K., Zhang, Y., Huang, J., Wang, L., et al. (2022). "SETMIL: spatial encoding transformer-based multiple instance learning for pathological image analysis," in *Medical image computing and computer assisted intervention – miccai 2022* (Cham: Springer Nature Switzerland), 66–76. doi:10.1007/978-3-031-16434-7_7

Zheng, Yi, Gindra, R. H., Green, E. J., Burks, E. J., Betke, M., Beane, J. E., et al. (2022). A graph-transformer for whole slide image classification. *IEEE Trans. Med. Imaging* 41 (11), 3003–3015. doi:10.1109/TMI.2022.3176598

Zhou, J., Chen, W., Wang, H., Shen, W., Xie, C., Yuille, A., et al. (2022). iBOT: image BERT pre-training with online tokenizer. *arXiv*. doi:10.48550/arXiv.2111.07832

Zhu, E., Qi, X., Huang, X., and Zhang, Z. (2024). Application of spatial omics in gastric cancer. *Pathology - Res. Pract.* 262 (October), 155503. doi:10.1016/j.prp.2024.155503

Zhu, W., Xie, L., Han, J., and Guo, X. (2020). The application of deep learning in cancer prognosis prediction. *Cancers* 12 (3), 603. doi:10.3390/cancers12030603

Zhuang, L., Park, S. H., Skates, S. J., Prosper, A. E., Aberle, D. R., and Hsu, W. (2025). Advancing precision oncology through modeling of longitudinal and multimodal data. *arXiv*, arXiv:2502.07836v3. doi:10.48550/arXiv.2502.07836

Zuo, Y., Wu, Y., Lu, Z., Qi, Z., Huang, K., Zhang, D., et al. (2022). "Identify consistent imaging genomic biomarkers for characterizing the survival-associated interactions between tumor-infiltrating lymphocytes and tumors," in *Medical image computing and computer assisted intervention – miccai 2022* (Cham: Springer Nature Switzerland), 222–231. doi:10.1007/978-3-031-16434-7_22

Żydowicz, W. M., Skokowski, J., Marano, L., and Polom, K. (2024). Current trends and beyond conventional approaches: advancements in breast cancer surgery through three-dimensional imaging, virtual reality, augmented reality, and the emerging metaverse. *J. Clin. Med.* 13 (3), 915. doi:10.3390/jcm13030915

Glossary

AI	Artificial intelligence
EHR	Electronic health record
MRI	Magnetic resonance imaging
CT	Computed tomography
TCGA	The Cancer Genome Atlas
PCA	Principal component analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
H&E	Hematoxylin and eosin
WSI	Whole-slide image
DL	Deep learning
SSL	Self-supervised learning
MLP	Multilayer perceptron
BYOL	Bootstrap Your Own Latent
DINO	Distillation with No Labels
iBOT	Image BERT Pre-Training with Online Tokenizer
MIL	Multi-instance learning
ViT	Vision Transformer
HIPT	Hierarchical Image Pyramid Transformer
CNN	Convolutional neural networks
NLP	Natural language processing
MSE	Mean square error
VQA	Visual Question Answering
ML	Machine learning
AIPS	AI-based predictive system
TMB	Tumor mutation burden
MSI	Microsatellite instable
AUC	Area under the curve
SALMON	survival analysis learning with multi-omics neural network
MTB	multidisciplinary tumor boards
RPA	Radiation Planning Assistant
AR	Augmented reality
VR	Virtual reality
SNN	self-normalizing neural network
SHAP	Shapley Additive Explanation
GDC	Genomic Data Commons
dbGAP	Database of Genotypes and Phenotypes
AACR project GENIE	American Association for Cancer Research project Genomics Evidence Neoplasia Information Exchange
TCIA	The Cancer Imaging Archive
EGA	European Genome-phenome Archive
GPIC	Genomics Pathology Imaging Collection
CPTAC	Clinical Proteomic Tumor Analysis Consortium
RNN	Recurrent neural network
RL	Reinforcement learning
NLST	National Lung Screening Trial
LGG	Brain Lower Grade Glioma
BRCA	Breast invasive carcinoma
COAD	Colon adenocarcinoma
GBM	Glioblastoma
BLCA	Bladder Urothelial Carcinoma
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
STAD	Stomach adenocarcinoma
UCEC	Uterine Corpus Endometrial Carcinoma
PAAD	Pancreatic adenocarcinoma
HNSC	Head and Neck squamous cell carcinoma.