Diagn Interv Radiol 2024; DOI: 10.4274/dir.2024.242631

ARTIFICIAL INTELLIGENCE AND INFORMATICS

INVITED REVIEW

The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review

Benjamin D. Simon[1,2]
Kutsev Bengisu Ozyoruk[1]
David G. Gelikman[1]
Stephanie A. Harmon[1]
Barış Türkbey[1]

[1]Molecular Imaging Branch, National Cancer Institute, National Institutes of Health, Bethesda, USA

[2]University of Oxford, Institute of Biomedical Engineering, Department Engineering Science, Oxford, UK

ABSTRACT

With the ongoing revolution of artificial intelligence (AI) in medicine, the impact of AI in radiology is more pronounced than ever. An increasing number of technical and clinical AI-focused studies are published each day. As these tools inevitably affect patient care and physician practices, it is crucial that radiologists become more familiar with the leading strategies and underlying principles of AI. Multimodal AI models can combine both imaging and clinical metadata and are quickly becoming a popular approach that is being integrated into the medical ecosystem. This narrative review covers major concepts of multimodal AI through the lens of recent literature. We discuss emerging frameworks, including graph neural networks, which allow for explicit learning from non-Euclidean relationships, and transformers, which allow for parallel computation that scales, highlighting existing literature and advocating for a focus on emerging architectures. We also identify key pitfalls in current studies, including issues with taxonomy, data scarcity, and bias. By informing radiologists and biomedical AI experts about existing practices and challenges, we hope to guide the next wave of imaging-based multimodal AI research.

KEYWORDS

Artificial intelligence, cancer research, multimodal, pathology, radiology

Artificial Intelligence (AI) is revolutionizing everyday life with its advanced capabilities in image processing, textual analysis, and more. Though this technology has only recently gained widespread public attention, its origins are not new. Research into neural networks began in the early to mid-20th century,[1] making it surprising that mainstream models, such as ChatGPT, which are now frequently cited in scientific literature, have only recently captured public interest.[2] Comparable to the emergence of computers in the 1940s, modern AI possesses a long-standing mathematical foundation but is still in its infancy.

The field of radiology is data-heavy, signal-rich, and technology-focused, making it a prime target for building AI applications. Thus, it is crucial that radiologists stay informed about methodological and clinical trends in AI. Radiologists routinely review large amounts of signal-rich data in a multimodal manner, making them well-suited to leverage AI and medical data to enhance diagnostic accuracy. At its core, AI is an extremely thorough pattern-detection system, capable of recognizing patterns beyond human capability for certain tasks. In medical imaging, which is nowadays very commonly used and results in work overload for practicing radiologists, AI has the potential to be a robust support tool within the radiology medical ecosystem. However, the introduction of AI raises ethical dilemmas[3] and security concerns,[4] including data leakage, automated medical decisions, biased data, and clinical impact.

While there is a growing body of literature on biomedical AI, much remains unexplored, particularly in the translation to medical applications. There has been a noticeable shift towards multimodal algorithms that incorporate imaging data with at least one other modality. Nevertheless, literature leveraging multimodal imaging data and clinical co-variates remains relatively sparse. For this reason, existing reviews on the topic have generally focused on 1) unimodal AI for imaging alone[5-7] or 2) general multimodal deep learning, which is becoming an increasingly heterogeneous field.[8-10] This review aims to explore multimodal AI in radiolo-

Corresponding author: Barış Türkbey

E-mail: turkbeyi@mail.nih.gov

Received 17 June 2024; revision requested 17 July 2024; accepted 04 September 2024.

Epub: 01.10.2024

Publication date: xx.xx.2024

DOI: 10.4274/dir.2024.242631

You may cite this article as: Simon BD, Ozyoruk KB, Gelikman DG, Harmon SA, Türkbey B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. *Diagn Interv Radiol.* 01 October 2024 DOI: 10.4274/dir.2024.242631 [Epub Ahead of Print].

gy comprehensively by examining both imaging and clinical variables. Throughout, we assess the methodology and clinical translation to inform future directions and organize approaches within the field.

Modern frameworks and multi-modality fusion techniques

The first focus of this study is the cutting-edge methodologies for multimodal AI. These frameworks are increasingly recognized as impactful approaches in advancing healthcare analytics due to their ability to interpret and integrate disparate forms of medical data, similar to the daily tasks of physicians. For detailed definitions and explanations of key terminology, a glossary of key terms with definitions is provided (Table 1). Central frameworks aim to model the relationship between data and corresponding clinical outcomes. Transformer-based models and graph neural networks (GNNs) have demonstrated remarkable promise in combining clinical notes,[11-13] imaging data,[14-16] and genomic information,[17-20] enhancing patient care through personalized and precise predictions and recommendations (Figure 1).

Transformers

Initially conceived for natural language processing, transformers have been adapted for other unimodal input data, such as imaging and genomics, and now, for multimodal tasks in healthcare. These models uniquely focus on different data components as needed and are adept at handling sequential data.[21] They also employ self-attention mechanisms, allowing for the assignment of weighted importance to different parts of input data, regardless of order. This implementation is especially beneficial for free text or genomic sequencing data, where the significance of a feature greatly depends on its context. These mechanisms have been extended to consider temporal dependencies in electronic health records (EHRs), enabling the model to discern which historical medical events are most predictive of future outcomes.[22]

Transformers are particularly revolutionary, unlike typical recurrent neural networks, in that they employ a parallelized approach, which allows for scalable computation.[23] Recurrent neural networks are a popular type of model that handle information sequentially and cannot do so in parallel.[24] This founda-

Table 1. Glossary of key terminology

Term	Our definition
Multimodal AI	AI models that integrate multiple types of data (e.g., imaging, clinical notes, genomic data) to improve diagnostic accuracy and patient outcomes.
Multichannel AI	AI models that integrate multiple inputs of the same type of data (e.g., multiple pathology images, multiple radiology images, multiple genomic sequences).
GNN	A type of neural network designed to capture dependencies in data that is structured as graphs, useful in settings where data interactions are non-linear and complex.
Transformers	A model architecture initially developed for natural language processing that has been adapted for analyzing various types of data. Known for its self-attention mechanism, which helps in understanding the importance of different parts of the data.
Machine learning	A method and field in computer science where systems are able to learn without deliberate instructions through mathematical pattern recognition of data.
AI	A broad field describing computer systems which are able to behave in ways that would normally require human intelligence.
Fusion techniques	Methods used to integrate multiple types of data in AI models. These can be early, joint, or late fusion, depending on when data types are combined during the model training process. There are many other statistical integration methods.
Parallel computation	A strategy in computer science where multiple processes or calculations happen simultaneously rather than one at a time.
Non-euclidean	Data that does not fit into traditional Euclidean geometry frameworks, such as graph data, which is essential for certain types of medical analyses where relationships and connections define data structure.
Clinical metadata	Information accompanying medical data that provides context about the health status, treatment, or diagnostics of a patient, crucial for interpreting imaging data in AI models.
Data curation	The process of organizing, integrating, and managing data collected from various sources to ensure it is accurate, complete, and reliable for AI training and analysis.
Self-attention mechanism	A component of neural network architectures that allows the model to weigh the importance of different parts of the input data differently, improving its ability to understand complex patterns.
Sequential data processing	In AI, the handling of data that is organized in a sequence (such as time series data from patient records), which is critical for understanding temporal patterns and dependencies.
Bias mitigation	Strategies and methodologies aimed at reducing bias in AI models to ensure fairness and equity, particularly important in healthcare applications where biased decisions can have serious implications.
Transfer learning	A machine learning method where a model developed for one task is reused as the starting point for a model on a second task, facilitating rapid deployment and reducing the need for large amounts of data.
Model generalizability	The ability of an AI model to perform well across different settings or populations, not just the ones on which it was trained, which is crucial for applications in diverse clinical environments.

AI, artificial intelligence; GNN, graph neural network.

Main points

- As multimodal artificial intelligence (AI) becomes increasingly integrated into the field of radiology, it is imperative that radiologists become familiar with the existing frameworks, applications, and analyses of such tools.
- Conventional approaches to multimodal AI integration have shown improvement over unimodal approaches in their ability to translate accurately to the clinic.
- Cutting-edge approaches for multimodal biomedical AI applications, such as transformers and graph neural networks, can integrate time series and non-Euclidean biomedical data.
- Key pitfalls of the multimodal biomedical AI landscape include inconsistent taxonomy, a lack of foundational models using varied large-scale representative data sources, and a mismatch between the healthcare arena and the necessary curation of data for AI models.

tional difference has led to transformers being the basis for large language models, such as BERT[25] and ChatGPT, but their application in medicine remains largely unexplored.[25,26] Literature using transformer-based multimodal predictions consistently finds that transformer models outperform typical recurrent or unimodal models.[27-30]

Despite the success of transformers, most literature features single-case applications, where a particular transformer architecture is optimized for a single clinical outcome.[31] A good example of an impactful application of transformers by Yu et al.[32] presents a framework to learn from imaging, clinical, and genetic information to set a new benchmark for diagnosing Alzheimer's disease (area under the receiver operator characteristic curve of 0.993). This work shows how transformers may be able to aid in unifying information across modalities for comprehensive learning in a specific disease space.

The literature on their broader optimization for various clinical or radiology tasks is limited. Khader et al.[33] propose a transferrable large-scale transformer approach, showing that it outperforms existing multimodal approaches leveraging convolutional neural networks (CNNs). They attribute their improvement to a novel technical approach, which selectively limits interactions between data inputs. They demonstrate the generalizability of their model by showing improvement across various decisions, including heart failure and respiratory disease prediction, and domains, including fundoscopy images and chest radiographs paired with non-imaging data.[33]

With the increasing popularity of multimodal data and models, there is a need for technical approaches that are transferrable and widely applicable for clinical use.

Graph neural networks

Although transformer-based models excel at capturing dependencies in sequential data,[34] their architecture does not inherently account for non-Euclidean structures present in multimodal healthcare data.[23] This gap has led to significant interest in GNNs, which model the data in a graph-structured format. This is particularly relevant to multimodal imaging data, where the relationships and dependencies between data points, such as between an anatomical structure in imaging and a genetic marker or clinical parameters, are not inherently grid-like and could be more accurately represented by graphical connections (Figure 2).

GNNs extend the concept of convolution from regular grids to graphs, with convolutional operations that aggregate feature information from a node's neighbors.[35] This approach captures global structural information. Unlike CNNs, where the same filter is applied uniformly across an image or matrix, GNNs adaptively learn how to weight the influence of neighboring nodes, making them adept at handling irregular data that does not conform to a fixed grid.[36]

This novelty is rooted in the ability of GNNs to learn from non-Euclidean data, which is crucial for integrating different types of medical information.[37] They can explicitly model the complex relationships between modalities, rather than attempting to map them in grid-like structures, such as CNNs, which may not fully take the structure into account[38] and could introduce biases related to artificial adjacency in grid formatting. Although exciting work has been taking place recently in medical imaging with GNNs, the bulk of multimodal literature continues to focus on CNNs, requiring tabular fusion in many cases.[39] There are several methodologies for fusing modalities.[40] However, without a graphical approach, there is potential for misinterpretation of the data's relationship when arbitrarily fused in a tabular format. For example, appending an image with a clinical parameter could falsely imply that parameters are adjacent to the imaging features. In contrast, with a GNN, this relationship can be modeled via nodes in a graphical representation, rather than being appended.

Despite the potential and applicability of GNNs, literature leveraging them in the medical space is scarce, likely due to their novelty and the varying custom methods for graphical construction posing a challenge. One study in the oncologic radiology space used a GNN to predict regional lymph node metastasis in esophageal squamous cell carcinoma patients.[41] In their work, Ding et al.[41] constructed a graph by mapping learned embeddings across image features and clinical parameters into a feature space, treating them each as nodes. They then used a graphical attention mechanism to learn the weights of the edges connecting the nodes. In another study, Gao et al.[20] used a completely different method for construction to predict the survival of cancer patients using gene expression data. They constructed a graph by considering each patient's primary modality encoding (which could be imaging, though they did not use imaging) as a node, with each gene also as a node. Edge weights were then determined by the level of gene expression for each patient and connected to the primary nodes. In a third study, Lyu et al.[42] demonstrate a successful GNN for predicting drug interactions by building graphs drawing edges between drugs and drug-related entities (such as targets or transporters). These three examples illustrate the complexity of graph construction and the custom nature of GNN methodology, which may explain the scarcity of literature on the topic despite its promise for relating multimodal data and encodings.

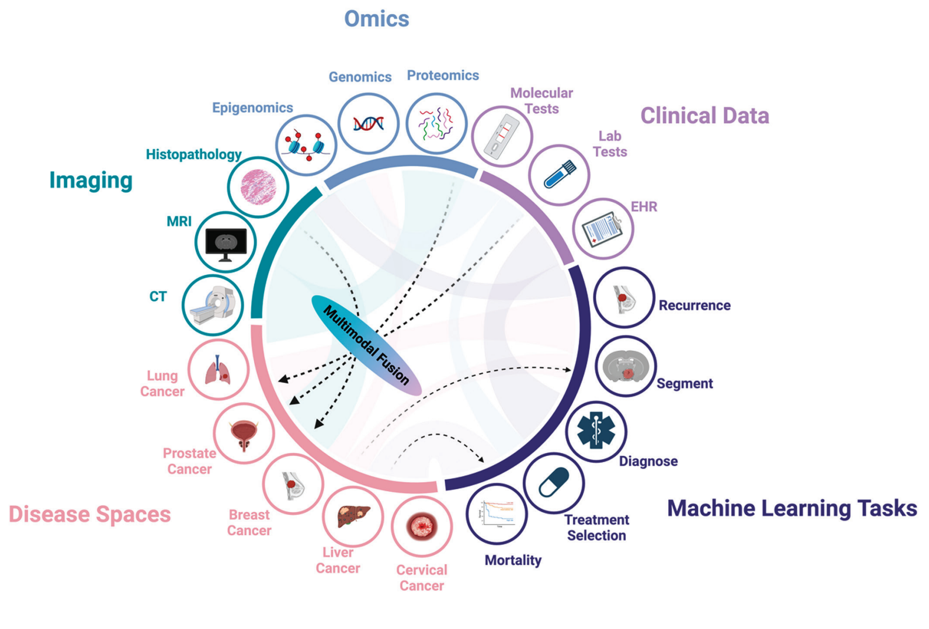

Figure 1. Multimodal medical artificial intelligence (AI) applications across disease spaces. Simplified schematic of the many applications of multimodal medical AI fusing imaging, omics, and clinical data for various tasks across disease spaces.

Modality fusion techniques

Despite the emergence of architectures such as GNNs, which can more deliberately represent data interactions, almost all medical data, whether imaging, molecular, or other signals, can be tabulated. Thus, various fusion techniques (methods for concatenating signals or information) are far more commonly used in multimodal literature.[9] Fusion techniques can broadly be categorized as early, intermediate/joint, or late fusion. In simple terms, early fusion means that the information is combined before learning via AI occurs, joint fusion means some learning happens before and after combining the two modalities, and late fusion means no learning happens after combining information. Therefore, it can be considered that late fusion aggregates learned information from the two modalities to make a prediction, whereas joint fusion allows for the modalities to interact, and for components of each to have complex relationships in making a prediction. More technically, early fusion generally involves concatenating input modalities into a single vector before feeding them into a model for training. These input modalities can be extracted features or raw data. Joint or intermediate fusion involves concatenating independently learned features prior to further learning. Late fusion generally refers to complete or almost complete learning occurring independently before concatenating vectors for a final activation and prediction. There has also been an emergence of "sketch" fusion, which is similar to early fusion, but rather than concatenation, modalities are translated to a common space. Schematics of early, joint, and late fusion pipelines are presented in Figure 3.

There is a rich and growing base of multimodal models using fusion to combine tabulated free speech,[43] genomic,[44,45] or clinical covariate data with images for diagnostics. Kumar et al.[43] combined X-ray images with audio data consisting of respiratory sounds and coughs for the diagnosis of coronavirus disease 2019. As a result, they showed that early detection is possible with 98.91% accuracy by fusing chest X-ray and cough models. There is limited consensus on the optimal fusion technique, perhaps due to variations in dataset quality, interactions between data sources, or the learning architectures. With many variables at play, developing a comprehensive approach to machine learning fusion, even for a single data type or disease case, becomes challenging. Each fusion modality may have advantages or disadvantages depending on the application, data set, and model architecture. Often, the best approach is to try all three and compare results. Conceptually, however, the pros and cons primarily depend on the concept of confounding variables. Consider the example of a hypothetical model for lung cancer outcome prediction where there are two modalities, one being clinical risk factors, such as cigarette consumption and obesity, and the second being genomic data. If these two modalities are believed to be additive and independent (non-confounding), the requirement may be for the AI to learn from them separately. In this case, late fusion may be appropriate. If it is believed there is significant crosstalk between the variables (the relationship between them is confounding), early or joint fusion may be more appropriate. Early fusion may be more appropriate when using smaller-scale genomic variant data that checks for a set of known variants that increase risk. Conversely, joint fusion may be more appropriate if the model is expected to learn variants of risk from a large amount of genomic sequencing data. Regardless, it is difficult to determine the optimal fusion strategy from the data alone and often worth exploring multiple approaches.

Although early fusion appears to be the most common fusion type across a variety of fields using imaging or imaging features combined with other modalities,[9,46-51] there are also numerous studies using joint[52-54] and late fusion.[55] The optimal fusion technique likely depends on the data source, architecture, and other specifics, making consensus challenging. It is important that researchers explore multiple fusion options when designing a multimodal model because, unfortunately, there are no guidelines for multimodal data fusion at this point in the field's development.

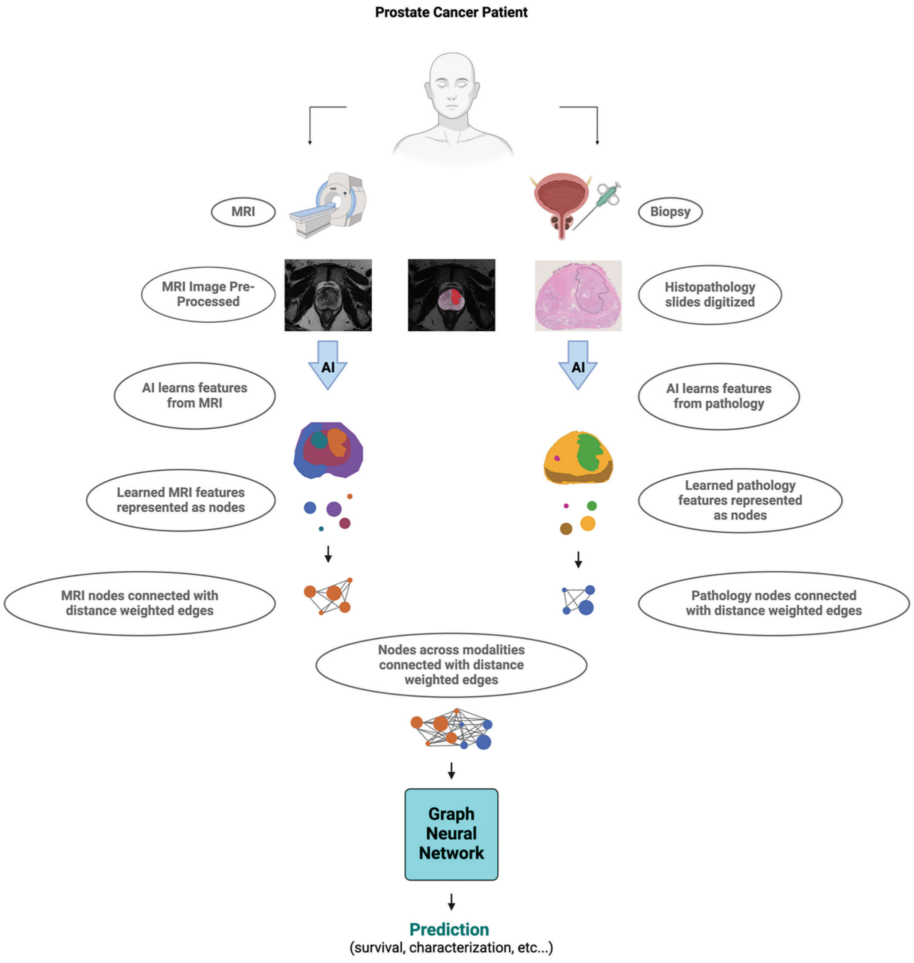
Prostate Cancer Patient
MRI
Biopsy
MRI Image Pre-Processed
Histopathology slides digitized
AI
AI
AI learns features from MRI
AI learns features from pathology
Learned MRI features represented as nodes
Learned pathology features represented as nodes
MRI nodes connected with distance weighted edges
Pathology nodes connected with distance weighted edges
Nodes across modalities connected with distance weighted edges
Graph Neural Network
Prediction
(survival, characterization, etc...)

Figure 2. Biomedical data for graph neural networks (GNNs). Example of a hypothetical application of a GNN in the prostate cancer space. Here, typical non-graph neural networks (labeled AI) learn features. Spatial relationships between these features of histopathology data and magnetic resonance images have the potential to be used in graph construction (using distance as the weights of edges and nodes corresponding to structures and features of pathology). AI, artificial intelligence.

In addition to these common concatenation techniques, there are many other examples of statistical integration methods. When it comes to GNNs, these integration methods can be customized to the relationship between specific modalities and datasets, as previously discussed. There are also many more methods outside the scope of this review, particularly pertaining to other omics data types. For example, mixOmics is a popular package for the integration and analysis of multi-omics data.[56] Other cutting-edge examples of multi-omics statistical integration frameworks include Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) and xMWAS.[57,58]

Current status of multimodal imaging work

The existing literature on multimodal AI contains numerous examples of successful multimodal integrations boasting impressive degrees of accuracy and proposed clinical translations.[59-69] These publications are promising and show the potential for multimodal AI implementation to improve patient outcomes. As the field progresses, there is an increase in highly curated large-scale data sets, paving the way for foundational models.[29,70] Nevertheless, much of the work in this space and its ability to translate to the clinic is limited by its siloed application, inconsistent taxonomy, and data scarcity.

Multimodal taxonomy

In the broad field of oncology, it is common for physicians to utilize multiple imaging channels to visualize abnormalities and make decisions. It follows that AI models leveraging multiple imaging sequences may be useful for tasks such as detection or segmentation. This raises the question: should combining two images be considered *multimodal*? Here, attention is drawn to the terms *multimodal* and *multichannel*. These terms are used in different and overlapping contexts across multiple disease spaces. In prostate cancer imaging literature, for example, the detection and segmentation of clinically significant prostate cancer are common goals often labeled as "multimodal" when merely integrating multiple magnetic resonance imaging (MRI) sequences, without incorporating fundamentally different data types.[71-75] Similar inconsistencies stand across the larger oncology field including, but not limited to, brain cancer,[76] lung cancer,[77,78] and breast cancer.[79]

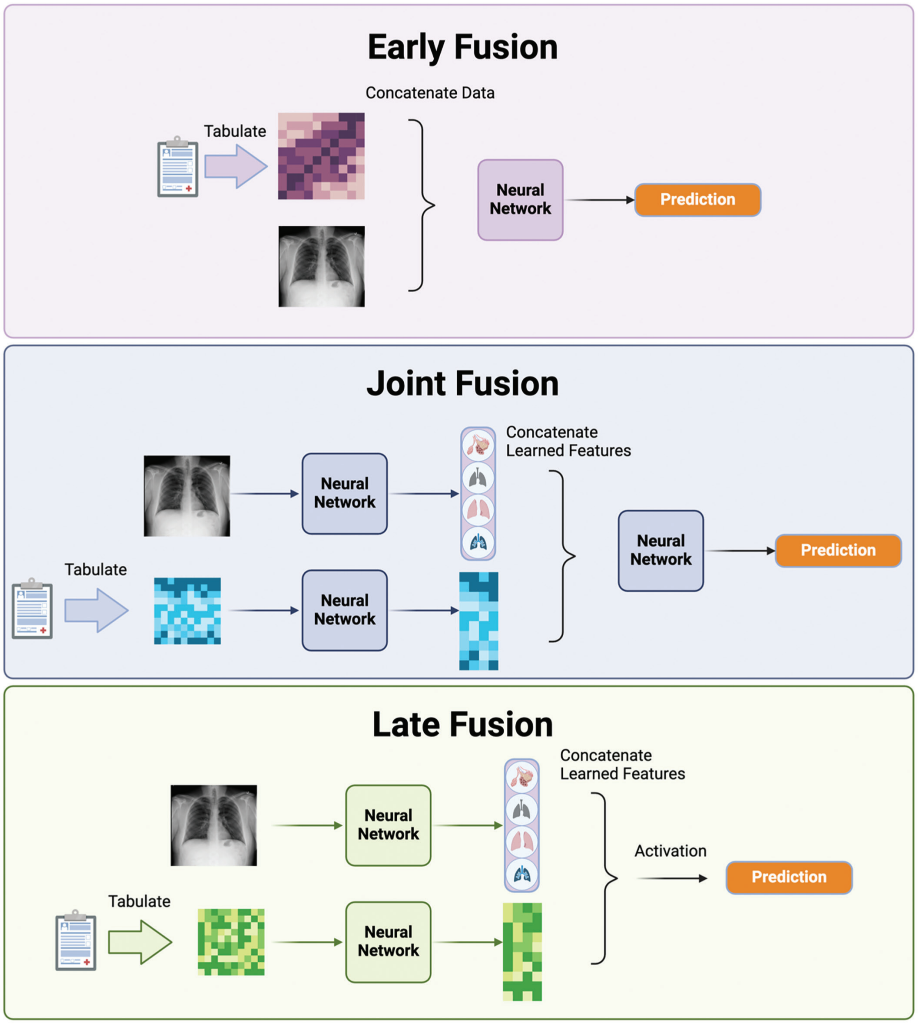

Figure 3. Fusion approaches for multimodal artificial intelligence (AI). Different approaches to training and fusion order are shown with examples in biomedical AI. In early fusion, a lung magnetic resonance image and tabulated electronic health record data are fused before learning. In joint fusion, one or both modalities undergo some learning prior to further learning and prediction. Finally, in late fusion, both modalities undergo all or almost all learning prior to fusion, activation, and prediction. There are many statistical integration methods beyond concatenation not shown.

The authors suggest that a multimodal model should combine conceptually different modes of information, whereas multichannel may be more appropriate for technically different (but categorically equivalent or similar) modes, as would be the case in fusing two radiologic images, such as multiple MRI sequences or computed tomography (CT) and MRI. Using this loose idea of "conceptually different images", one may consider combining digital histopathology images with radiomics as multimodal,[80] but the examples above (of fusing two radiologic images) would likely be considered multichannel and unimodal. In the authors' work with deep learning in the prostate cancer space, these image fusion models have been referred to as multichannel rather than multimodal.[81,82] With this pattern being evident across disease spaces, there is a need to clarify the taxonomy as the term "multimodal" becomes increasingly imprecise.

Generalizable models with transferrable application

The multimodal AI space is rapidly expanding but remains ultra-specific, hindering the transition of findings into general practices. Building models that translate across regions and hospitals without bias may be better explored through foundational models that 1) apply to multiple disease spaces, 2) inform future methodological decision-making by outlining the evidence for engineering decisions or by demonstrating that a method is effective beyond a single isolated case, and 3) prove multicenter validation for clinical use with resistance to bias.

This trend is becoming apparent as the unimodal clinical AI space becomes increasingly saturated, and the most impactful publications focus on foundational models through novel technical innovations, such as with DINO,[83] DINOv2,[84] and iBOT,[85] increasingly large datasets, and self-supervised learning to leverage unannotated data.[86] This generalizability has yet to become commonplace in multimodal AI, except for some key examples. For instance, Khader et al.[29] provide a compelling case for multimodal transformers by analyzing 25 conditions using imaging and non-imaging patient data from the Medical Information Mart for Intensive Care (MIMIC), instead of evaluating a single disease case. This publication is an impressive example of using up-to-date methodologies (namely, transformers), baseline comparison to alternate approaches for the same dataset, and analysis of various conditions. They observed improvement through multimodal use across all disease cases and reported appropriate statistical evaluation. Unfortunately, it is not common practice for multimodal papers to present statistics compared with a baseline unimodal model or present evidence of value in including both modalities. Rather, such papers often present a means to an end. Instead, Khader et al.[29] provided a case for a specific method, informing how future researchers should proceed while highlighting multiple translational impacts.

Another example study pushing towards generalizable multimodal approaches is proposed by Soenksen et al.[70] They propose and assess a model for Holistic AI in Medicine (HAIM) to support the general development and testing of a variety of multimodal AI systems. Leveraging the MIMIC database, they demonstrate improvement in predicting various healthcare operations including lung lesion detection, 48-hour mortality, and edema. They find that all multimodal inputs improve performance across all predictions. However, there is no statistical analysis presented to inform us which of these tasks shows a statistically significant difference. This work pushes the medical field towards cutting-edge and generalizable multimodal work and emphasizes the need to develop a standard of comparison in the field.[70]

It is noteworthy, but not coincidental, that both models discussed above leverage the same MIMIC database. The MIMIC database is a publicly available repository of EHRs from the Beth Israel Deaconess Medical Center.[87,88] Though each publication attempts to draw data from multiple sources, this highlights the issue of database bias in designing multimodal algorithms.

Dataset curation

Database bias can manifest in various ways. Based on an analysis of the existing terrain, several examples of bias where the field may be at risk are discussed. As other reviews[8,89] and even the original MIMIC-IV publication[88] have stated, data in hospitals today is typically stored in systems not conducive to or able to support research, especially data science research. Built for security and far behind modern standards for user interface design, storage, and ease of access, it is not uncommon to find scanned versions of electronic medical records as PDF-format files, equivalent information stored in various locations at different hospitals, and logging methods varying between physicians. In other words, there is a significant mismatch between the data format resulting from existing data collection practices across healthcare facilities and the data format necessary for appropriate AI development. These mismatches make it quite challenging to curate datasets such as MIMIC, which require careful planning, financial investment, and an industry-wide shift in how medical data is collected and stored. As a result, models are at risk of being overtrained on the limited existing AI-friendly data.

By using a single center or focusing on training with the handful of carefully curated datasets available, models can "learn" to treat all patients as they would in those specific settings and time periods, regardless of the quality of care one receives at their own institution and the clinical environment of which they are a part. Clinical outcomes can vary significantly depending on the surgeon, environmental exposures, or technology available. For example, patients at the best hospitals in the country may have different outcomes from average hospitals and therefore should be treated differently. Beyond social determinants of health, from a technical perspective, considering that MRI or CT scanners may differ across the country, a model may inadvertently learn that image quality is associated with outcomes or be unable to accurately assess certain images. As with comparing baseline unimodal models, there is a need for guidelines to assess and mitigate bias in AI as it becomes more widespread. Although there are examples of papers identifying or discussing bias,[90-93] few propose analytical frameworks to address or measure bias in AI.[94] Such publications are varied, and none have become standard practice in the field. Few clinical papers assess bias in clinically specific AI models. Though not multimodal, a machine learning approach was proposed by Chandran et al.[95] to predict lung cancer risk using the cross-area under the receiver operator characteristic curve to measure disparities in performance by race and ethnicity. They identify key failures in the model's ability to determine risk for Asian and Hispanic individuals compared with White and non-Hispanic individuals. The mismatch between the clinical environment and AI-friendly data storage requirements results not only in bias but also makes bias reduction challenging, as curating "representative" data from centers across the entire country is a huge undertaking. The more representative the training data is of the setting in which it is applied, the lower the risk of biased decisions. With evidence that multimodal AI may be more accurate for some AI applications[27,59-69,80,93] and that multimodal work is more challenging to curate consistently across institutions, researchers and physicians face the decision of how to build and employ AI tools when smaller multimodal sample sizes promise improved overall accuracy, but smaller sample size may increase risk of bias.

This concern is currently pressing and needs to be addressed. One systematic review on GNNs based on EHRs reported that out of 50 papers reviewed, 23 used MIMIC-III and 6 MIMIC-IV.[96] With the increasing prevalence of AI research and rapid translation of tools to the clinic, there is a need for a change in how data is stored and collected by healthcare providers across the country. Continuing to develop AI tools on the available pool of high-quality curated datasets, such as MIMIC,[87,88] the UK Biobank,[97] EMBED,[98] and the Scottish Medical Imaging Archive,[99] is risky as tools may be carelessly applied to populations with differing clinical environments or health outcomes. Further, with the medical field being a rapidly changing ecosystem, models and datasets can quickly become less relevant to the current medical system.

Considering the dynamic medical environment and its quickly changing technology and guidelines, AI and the data on which it is trained will have to change as quickly as the clinic. One must be incredibly mindful of the dynamic nature of data when training an AI algorithm. Here, "dynamic" can take on a double meaning. Data can be dynamic in that its surrounding clinical environment changes as knowledge and technology develop. It can also be dynamic in that the information itself changes as a product of aging or biological changes. For example, considering genomics are stable over time, it is unclear what the significance is of their integration with dynamic data, such as an imaging phenotype or proteomics, which can change over a person's life. Imaging data or radiomics data have been integrated both with stable omics[100] and dynamic omics for multimodal AI.[101] Regardless of the biomedical data and if it is dynamic, a change in how the data is collected from, and impacts, the clinic may be just as impactful on the creation of impactful AI as the data itself.

AI tools have the potential to both combat and exacerbate biases by providing evidence-based recommendations. Radiologists and other physicians must understand emerging and existing methods in the field, as well as the importance of data set curation, as they are often the ones making final decisions about how these tools will be used and how they will impact the patient. By being aware of the potential for AI to exacerbate biases, radiologists are relied upon to view these tools as exactly what they are: physician *support* tools. Even if a tool has a proven record of being more accurate than the average physician at, for example, detecting lesions on a certain type of scan, there will still be mistakes, and physicians will need to be able to use these AI tools without catering to their biases. It is difficult to predict exactly what the role of radiologists will be in the future of using and developing AI, but the reality is that it will play a role. The greater the degree to which these tools are understood is, the easier it will be for physicians to interact with them in a way that improves health. On the flip side, a greater understanding among physicians will allow them to conduct their clinic in a way that is conducive to storing data for training strong bias-mitigated models.

Future directions

Multimodal AI will inevitably continue to develop and be explored through the methodologies, foundational models, and translational integrations discussed in this review and beyond. Despite exploring highly developed architectures, methods, and techniques in image processing AI, such as fusion models, transformers, and GNNs, the medical field lags in using up-to-date AI innovations and struggles with consistency in taxonomy, evaluation metrics, and methodology, even within the same disease spaces.

The lack of common practices, which will develop and change as the field matures, severely limits progress and translation. It becomes difficult to generalize conclusions from one publication to the next and across methodologies. Standout publications in the multimodal AI space are characterized by their ability to generalize as foundational models with transferrable applications, incorporate physician perspectives with clear and broad clinical utility, and carefully evaluate baseline models using thorough and appropriate evaluation and statistics.

An even more pressing limitation in developing multimodal AI tools with biomedical applications is the lack of comprehensive, high-quality data. As discussed, most reviewed works rely on either a very small set of carefully curated data, which requires extensive time, resources, and funding for AI development, or they draw from a select set of high-quality, open-access datasets. By repeatedly using these same high-quality curated datasets, a suite of AI-based translational tools heavily biased toward the included locations, periods, and patient populations is being developed. With the clinical setting and its outcomes being a constantly changing ecosystem, it is risky to rely on the same datasets. Equitable, bias-free AI will require these systems to be dynamic, constantly updated with new data, and capable of adapting over time with fine-tuning. Technologists and clinicians may have to meet somewhere in the middle, such that technologists will have to build models using less-than-optimal data, and clinicians may have to incorporate certain practices into their data ecosystem to ensure AI models are up to date.

Our narrative review of multimodal AI, combining imaging and other clinical metadata, aims to propose clarifications for what constitutes "multimodal" AI in imaging, identify up-to-date frameworks with potential for enhanced results in future model research, comment on a shift toward generalizable foundational models, and identify trends and concerns in database curation. As the field progresses from theory to clinic, it is essential for radiologists to stay informed about the latest developments, methodologies, and ethical implications.

The current radiologic landscape is characterized by a transition toward multimodal fusion models, with increasing focus on transformers and GNNs. However, there is a considerable amount of work to be done in terms of scientific due diligence regarding gaps in methodology and model training bias. Moreover, the reliance on the few existing high-quality curated datasets highlights a major risk as AI tools become more common in the clinical setting. There is an urgent need to align the format of data required for training AI with that logged by physicians to curate comprehensive training databases.

In conclusion, while AI in radiology promises significant advancements in the field, successful and unbiased integration demands a multidisciplinary approach involving continuous education of physicians and AI developers alike. By informing radiologists, we hope to begin bridging the gap between technology and the clinic, guiding future methodologies, practices for dataset curation, and the field as a whole. By harnessing the power of AI, appropriate evaluation, and physician expertise, we hope to save more lives and improve the quality of care for patients worldwide.

Conflict of interest disclosure

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Barış Türkbey receives support as part of a CRADA between NIH and NVIDIA and between NIH and Philips Healthcare, receives royalties from NIH, and is a party to patents or potential patents related to this work. The remaining authors declare that there are no other disclosures relevant to the subject matter of this article. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. All figures were created and licensed using BioRender.com.

References

1. Eberhart RC, Dobbins RW. Early neural network development history: the age of Camelot. *IEEE Eng Med Biol Mag*. 1990;9(3):15-18. [CrossRef]
2. Temsah MH, Altamimi I, Jamal A, Alhasan K, Al-Eyadhy A. ChatGPT Surpasses 1000 Publications on PubMed: Envisioning the Road Ahead. *Cureus*. 2023;15(9):e44769. [CrossRef]
3. Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the ethical enigma: artificial intelligence in healthcare. *Cureus*. 2023;15(8):e43262. [CrossRef]
4. Li J. Security implications of AI chatbots in health care. *J Med Internet Res*. 2023;25:e47551. [CrossRef]
5. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ Digit Med*. 2022;5(1):48. [CrossRef]
6. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health*. 2020;2(9):e486-e488. [CrossRef]
7. Esteva A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med*. 2021;4(1):5. [CrossRef]

8. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773-1784. [CrossRef]

9. Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. *NPJ Digit Med*. 2022;5(1):171. [CrossRef]

10. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. *Inf Fusion*. 2022;77:29-52. [CrossRef]

11. Lyu W, Dong X, Wong R, et al. A Multimodal transformer: fusing clinical notes with structured EHR data for interpretable in-hospital mortality prediction. *AMIA Annu Symp Proc*. 2022;2022:719-728. [CrossRef]

12. Rahman W, Hasan MK, Lee S, et al. Integrating multimodal information in large pretrained transformers. *Proc Conf Assoc Comput Linguist Meet*. 2020;2020:2359-2369. [CrossRef]

13. Rohanian O, Nouriborji M, Jauncey H, et al. Lightweight transformers for clinical natural language processing. *Nat Lang Eng*. 2024. [CrossRef]

14. Keicher M, Burwinkel H, Bani-Harouni D, et al. Multimodal graph attention network for COVID-19 outcome prediction. *Sci Rep*. 2023;13(1):19539. [CrossRef]

15. Zhou HY, Yu Y, Wang C, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat Biomed Eng*. 2023;7(6):743-755. [CrossRef]

16. Nguyen HH, Blaschko MB, Saarakkala S, Tiulpin A. Clinically-inspired multi-agent transformers for disease trajectory forecasting from multimodal data. *IEEE Trans Med Imaging*. 2024;43(1):529-541. [CrossRef]

17. Li B, Nabavi S. A multimodal graph neural network framework for cancer molecular subtype classification. *Bmc Bioinformatics*. 2024;25(1):27. [CrossRef]

18. Wen HZ, Ding JY, Jin W, Wang YQ, Xie YY, Tang JL. Graph neural networks for multimodal single-cell data integration. *Proceedings of the 28th Acm Sigkdd Conference on Knowledge Discovery and Data Mining, Kdd 2022*. 2022:4153-4163. [CrossRef]

19. Zhu JN, Oh JH, Simhal AK, et al. Geometric graph neural networks on multi-omics data to predict cancer survival outcomes. *Comput Biol Med*. 2023;163:107117. [CrossRef]

20. Gao J, Lyu T, Xiong F, Wang J, Ke W, Li Z. Predicting the survival of cancer patients with multimodal graph neural network. *IEEE/ACM Trans Comput Biol Bioinform*. 2022;19(2):699-709. [CrossRef]

21. Alam F, Ananbeh O, Malik KM, et al. towards predicting length of stay and identification of cohort risk factors using self-attention-based transformers and association mining: COVID-19 as a phenotype. *Diagnostics (Basel)*. 2023;13(10):1760. [CrossRef]

22. Yang Z, Mitra A, Liu W, Berlowitz D, Yu H. TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. *Nat Commun*. 2023;14(1):7857. [CrossRef]

23. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neur In*. 2017:30. [CrossRef]

24. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. *JMLR Workshop Conf Proc*. 2016;56:301-318. [CrossRef]

25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota. Association for Computational Linguistics;* 2019;1:4171-4186. [CrossRef]

26. Nazir A, Wang Z. A Comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges. *Meta Radiol*. 2023;1(2):100022. [CrossRef]

27. Khader F, Kather JN, Müller-Franzes G, et al. Medical transformer for multimodal survival prediction in intensive care: integration of imaging and non-imaging data. *Sci Rep*. 2023;13(1):10666. [CrossRef]

28. Zheng H, Lin Z, Zhou Q, et al. Multi-transSP: Multimodal transformer for survival prediction of nasopharyngeal carcinoma patients. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S, eds. Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science. Springer, Cham; 2022:234-243. [CrossRef]

29. Khader F, Müller-Franzes G, Wang TC, et al. Multimodal deep learning for integrating chest radiographs and clinical parameters: a case for transformers. *Radiology*. 2023;309(1):e230806. [CrossRef]

30. Li Y, Rao S, Solares JRA, et al. BEHRT: transformer for electronic health records. *Sci Rep*. 2020;10(1):7155. [CrossRef]

31. Liu L, Liu S, Zhang L, To XV, Nasrallah F, Chandra SS. Cascaded multi-modal mixing transformers for alzheimer's disease classification with incomplete data. *Neuroimage*. 2023;277:120267. [CrossRef]

32. Yu Q, Ma Q, Da L, et al. A transformer-based unified multimodal framework for Alzheimer's disease assessment. *Comput Biol Med*. 2024;180:108979. [CrossRef]

33. Khader F, Franzes GM, Wang T, et al. Medical diagnosis with large scale multimodal transformers: leveraging diverse data for more accurate diagnosis. 2022. [CrossRef]

34. Guo J, Jia N, Bai J. Transformer based on channel-spatial attention for accurate classification of scenes in remote sensing image. *Sci Rep*. 2022;12(1):15473. [CrossRef]

35. Meng X, Zou T. Clinical applications of graph neural networks in computational histopathology: a review. *Comput Biol Med*. 2023;164:107201. [CrossRef]

36. Khemani B, Patil S, Kotecha K, Tanwar S. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *J Big Data*. 2024;11(1):18. [CrossRef]

37. Zhou Y, Huo H, Hou Z, et al. Co-embedding of edges and nodes with deep graph convolutional neural networks. *Sci Rep*. 2023;13(1):16966. [CrossRef]

38. Zhang XM, Liang L, Liu L, Tang MJ. Graph neural networks and their current applications in bioinformatics. *Front Genet*. 2021;12:690049. [CrossRef]

39. Li MM, Huang K, Zitnik M. Graph representation learning in biomedicine and healthcare. *Nat Biomed Eng*. 2022;6(12):1353-1369. [CrossRef]

40. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med*. 2020;3:136. [CrossRef]

41. Ding M, Cui H, Li B, et al. Integrating preoperative computed tomography and clinical factors for lymph node metastasis prediction in esophageal squamous cell carcinoma by feature-wise attentional graph neural network. *Int J Radiat Oncol Biol Phys*. 2023;116(3):676-689. [CrossRef]

42. Lyu T, Gao J, Tian L, Li Z, Zhang P, Zhang J. MDNN: a multimodal deep neural network for predicting drug-drug interaction events. 2021:3536-3542. [CrossRef]

43. Kumar S, Chaube MK, Alsamhi SH, et al. A novel multimodal fusion framework for early diagnosis and accurate classification of COVID-19 patients using X-ray images and speech signal processing techniques. *Comput Methods Programs Biomed*. 2022;226:107109. [CrossRef]

44. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer*. 2022;3(10):1151-1164. [CrossRef]

45. Subramanian V, Do MN, Syeda-Mahmood T. Multimodal fusion of imaging and genomics for lung cancer recurrence prediction. 2020:arXiv:2002.01982. [CrossRef]

46. Cearns M, Opel N, Clark S, et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Transl Psychiatry*. 2019;9(1):285. [CrossRef]

47. Cheng B, Liu M, Suk HI, Shen D, Zhang D; Alzheimer's disease neuroimaging initiative. Multimodal manifold-regularized transfer

learning for MCI conversion prediction. *Brain Imaging Behav*. 2015;9(4):913-926. [CrossRef]

48. Peeken JC, Goldberg T, Pyka T, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med*. 2019;8(1):128-136. [CrossRef]

49. Zhou H, Chang K, Bai HX, et al. Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas. *J Neurooncol*. 2019;142(2):299-307. [CrossRef]

50. Brugnara G, Neuberger U, Mahmutoglu MA, et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke*. 2020;51(12):3541-3551. [CrossRef]

51. Qin H, Hu X, Zhang J, et al. Machine-learning radiomics to predict early recurrence in perihilar cholangiocarcinoma after curative resection. *Liver Int*. 2021;41(4):837-850. [CrossRef]

52. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. 7-point checklist and skin lesion classification using multi-task multi-modal neural nets. *IEEE J Biomed Health Inform*. 2018. [CrossRef]

53. Spasov SE, Passamonti L, Duggento A, Lio P, Toschi N. A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018;2018:1271-1274. [CrossRef]

54. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*. 2019;292(1):60-66. [CrossRef]

55. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol*. 2018;27(11):1261-1267. [CrossRef]

56. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13(11):e1005752. [CrossRef]

57. Singh A, Shannon CP, Gautier B, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics*. 2019;35(17):3055-3062. [CrossRef]

58. Uppal K, Ma C, Go YM, Jones DP, Wren J. xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics*. 2018;34(4):701-702. [CrossRef]

59. Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci Rep*. 2020;10(1):22147. [CrossRef]

60. Koutsouleris N, Dwyer DB, Degenhardt F, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry*. 2021;78(2):195-209. [CrossRef]

61. Rajpurkar P, Park A, Irvin J, et al. AppendiXNet: deep learning for diagnosis of appendicitis from a small dataset of CT exams using video pretraining. *Sci Rep*. 2020;10(1):3958. [CrossRef]

62. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med*. 2022;6(1):71. Erratum in: *NPJ Digit Med*. 2023;6(1):27. [CrossRef]

63. Janßen C, Boskamp T, Le'Clerc Arrastia J, et al. Multimodal lung cancer subtyping using deep learning neural networks on whole slide tissue images and MALDI MSI. *Cancers (Basel)*. 2022;14(24):6181. [CrossRef]

64. Khan RA, Fu M, Burbridge B, Luo Y, Wu FX. A multi-modal deep neural network for multi-class liver cancer diagnosis. *Neural Netw*. 2023;165:553-561. [CrossRef]

65. Steyaert S, Qiu YL, Zheng Y, Mukherjee P, Vogel H, Gevaert O. Multimodal deep learning to predict prognosis in adult and pediatric brain tumors. *Commun Med (Lond)*. 2023;3(1):44. [CrossRef]

66. Yao Y, Lv Y, Tong L, et al. ICSDA: a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data. *Brief Bioinform*. 2022;23(6):bbac448. [CrossRef]

67. Schulz S, Woerl AC, Jungmann F, et al. Multimodal deep learning for prognosis prediction in renal cancer. *Front Oncol*. 2021;11:788740. [CrossRef]

68. Boehm KM, Aherne EA, Ellenson L, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat Cancer*. 2022;3(6):723-733. [CrossRef]

69. Dayan I, Roth HR, Zhong A, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med*. 2021;27(10):1735-1743. [CrossRef]

70. Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit Med*. 2022;5(1):149. [CrossRef]

71. Nai YH, Teo BW, Tan NL, et al. Evaluation of multimodal algorithms for the segmentation of multiparametric MRI prostate images. *Comput Math Methods Med*. 2020;2020:8861035. [CrossRef]

72. Bleker J, Kwee TC, Dierckx RAJO, de Jong IJ, Huisman H, Yakar D. Multiparametric MRI and auto-fixed volume of interest-based radiomics signature for clinically significant peripheral zone prostate cancer. *Eur Radiol*. 2020;30(3):1313-1324. [CrossRef]

73. Chen Q, Xu X, Hu S, Li X, Zou Q, Li Y. A transfer learning approach for classification of clinical significant prostate cancers from mpMRI scans. *Proc SPIE Int Soc Opt Eng*. 2017:10134. [CrossRef]

74. Mehrtash A, Sedghi A, Ghafoorian M, et al. Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks. *Proc SPIE Int Soc Opt Eng*. 2017:10134. [CrossRef]

75. Gutierrez Y, Arevalo J, Martinez F. Multimodal contrastive supervised learning to classify clinical significance MRI regions on prostate cancer. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022:1682-1685. [CrossRef]

76. Nie D, Lu J, Zhang H, et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Sci Rep*. 2019;9(1):1103. [CrossRef]

77. Shim SO, Alkinani MH, Hussain L, Aziz W. Feature ranking importance from multimodal radiomic texture features using machine learning paradigm: a biomarker to predict the lung cancer. *Big Data Res*. 2022;29(28). [CrossRef]

78. Dong Y, Hou L, Yang W, et al. Multi-channel multi-task deep learning for predicting EGFR and KRAS mutations of non-small cell lung cancer on CT images. *Quant Imaging Med Surg*. 2021;11(6):2354-2375. [CrossRef]

79. Lu GX, Tian RH, Yang W, et al. Deep learning radiomics based on multimodal imaging for distinguishing benign and malignant breast tumours. *Front Med (Lausanne)*. 2024;11:1402967. [CrossRef]

80. Hu C, Qiao X, Huang R, Hu C, Bao J, Wang X. Development and validation of a multimodality model based on whole-slide imaging and biparametric MRI for predicting postoperative biochemical recurrence in prostate cancer. *Radiol Imaging Cancer*. 2024;6(3):e230143. [CrossRef]

81. Mehralivand S, Yang D, Harmon SA, et al. A cascaded deep learning-based artificial intelligence algorithm for automated lesion detection and classification on biparametric prostate magnetic resonance imaging. *Acad Radiol*. 2022;29(8):1159-1168. [CrossRef]

82. Simon BD, Merriman KM, Harmon SA, et al. Automated detection and grading of extraprostatic extension of prostate cancer at MRI via cascaded deep learning and random forest classification. *Acad Radiol*. 2024. [CrossRef]

83. Caron M, Touvron H, Misra I, et al. Emerging properties in self-supervised vision transformers. *2021 Ieee/Cvf International Conference on Computer Vision (Iccv 2021)*. 2021:9630-9640. [CrossRef]

84. Oquab M, Darcet T, Moutakanni T, et al. DINOv2: learning robust visual features without supervision. 2023:arXiv:2304.07193. Accessed April 01, 2023. [CrossRef]

85. Zhou J, Wei C, Wang H, et al. iBOT: image BERT pre-training with Online Tokenizer. 2021:arXiv:2111.07832. Accessed November 01, 2021. [CrossRef]

86. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-265. [CrossRef]

87. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035. [CrossRef]

88. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. [CrossRef]

89. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR data science institute artificial intelligence survey. *J Am Coll Radiol*. 2021;18(8):1153-1159. [CrossRef]

90. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. 2021;27(12):2176-2182. [CrossRef]

91. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. [CrossRef]

92. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc*. 2020;27(12):2020-2023. [CrossRef]

93. Kocak B, Ponsiglione A, Stanzione A, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagn Interv Radiol*. 2024. [CrossRef]

94. Yang J, Soltan AAS, Eyre DW, Yang Y, Clifton DA. An adversarial training framework for mitigating algorithmic biases in clinical machine learning. *NPJ Digit Med*. 2023;6(1):55. [CrossRef]

95. Chandran U, Reps J, Yang R, Vachani A, Maldonado F, Kalsekar I. Machine learning and real-world data to predict lung cancer risk in routine care. *Cancer Epidemiol Biomarkers Prev*. 2023;32(3):337-343. [CrossRef]

96. Oss Boll H, Amirahmadi A, Ghazani MM, et al. Graph neural networks for clinical risk prediction based on electronic health records: A survey. *J Biomed Inform*. 2024;151:104616. [CrossRef]

97. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. Mar 2015;12(3):e1001779. [CrossRef]

98. Jeong JJ, Vey BL, Bhimireddy A, et al. The EMory BrEast imaging Dataset (EMBED): A racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images. *Radiol Artif Intell*. 2023;5(1):e220047. [CrossRef]

99. Baxter R, Nind T, Sutherland J, et al. The scottish medical imaging archive: 57.3 million radiology studies linked to their medical records. *Radiol Artif Intell*. 2024;6(1):e220266. [CrossRef]

100. Huang YJ, Chen CH, Yang HC. AI-enhanced integration of genetic and medical imaging data for risk assessment of Type 2 diabetes. *Nat Commun*. 2024;15(1):4230. [CrossRef]

101. Lai J, Chen Z, Liu J, et al. A radiogenomic multimodal and whole-transcriptome sequencing for preoperative prediction of axillary lymph node metastasis and drug therapeutic response in breast cancer: a retrospective, machine learning and international multicohort study. *Int J Surg*. 2024;110(4):2162-2177. [CrossRef]