BJR https://doi.org/10.1259/bjr.20230211

Received: 28 February 2023 | Accepted: 27 June 2023 | Published online: 03 September 2023

Cite this article as:
Wei L, Niraula D, Gates EDH, Fu J, Luo Y, Nyflot MJ, et al. Artificial intelligence (AI) and machine learning (ML) in precision oncology: a review on enhancing discoverability through multiomics integration. *Br J Radiol* (2023) 10.1259/bjr.20230211.

AI IN IMAGING AND THERAPY: INNOVATIONS, ETHICS, AND IMPACT: REVIEW ARTICLE

Artificial intelligence (AI) and machine learning (ML) in precision oncology: a review on enhancing discoverability through multiomics integration

[1]LISE WEI, [2]DIPESH NIRAULA, [3]EVAN D. H. GATES, [4]JIE FU, [2]YI LUO, [3]MATTHEW J. NYFLOT, [3]STEPHEN R. BOWEN, [2]ISSAM M EL NAQA and [3]SUNAN CUI

[1]Department of Radiation Oncology, University of Michigan, Michigan, United States
[2]Department of Radiation Oncology, Moffitt Cancer Center, Tampa, United States
[3]Department of Radiation Oncology, University of Washington, Washington, United States
[4]Department of Radiation Oncology, Stanford University, Stanford, California, United States

Address correspondence to:
Lise Wei
E-mail: *liswei@umich.edu*
Sunan Cui
E-mail: *sunanc@uw.edu*

ABSTRACT

Multiomics data including imaging radiomics and various types of molecular biomarkers have been increasingly investigated for better diagnosis and therapy in the era of precision oncology. Artificial intelligence (AI) including machine learning (ML) and deep learning (DL) techniques combined with the exponential growth of multiomics data may have great potential to revolutionize cancer subtyping, risk stratification, prognostication, prediction and clinical decision-making.

In this article, we first present different categories of multiomics data and their roles in diagnosis and therapy. Second, AI-based data fusion methods and modeling methods as well as different validation schemes are illustrated. Third, the applications and examples of multiomics research in oncology are demonstrated. Finally, the challenges regarding the heterogeneity data set, availability of omics data, and validation of the research are discussed. The transition of multiomics research to real clinics still requires consistent efforts in standardizing omics data collection and analysis, building computational infrastructure for data sharing and storing, developing advanced methods to improve data fusion and interpretability, and ultimately, conducting large-scale prospective clinical trials to fill the gap between study findings and clinical benefits.

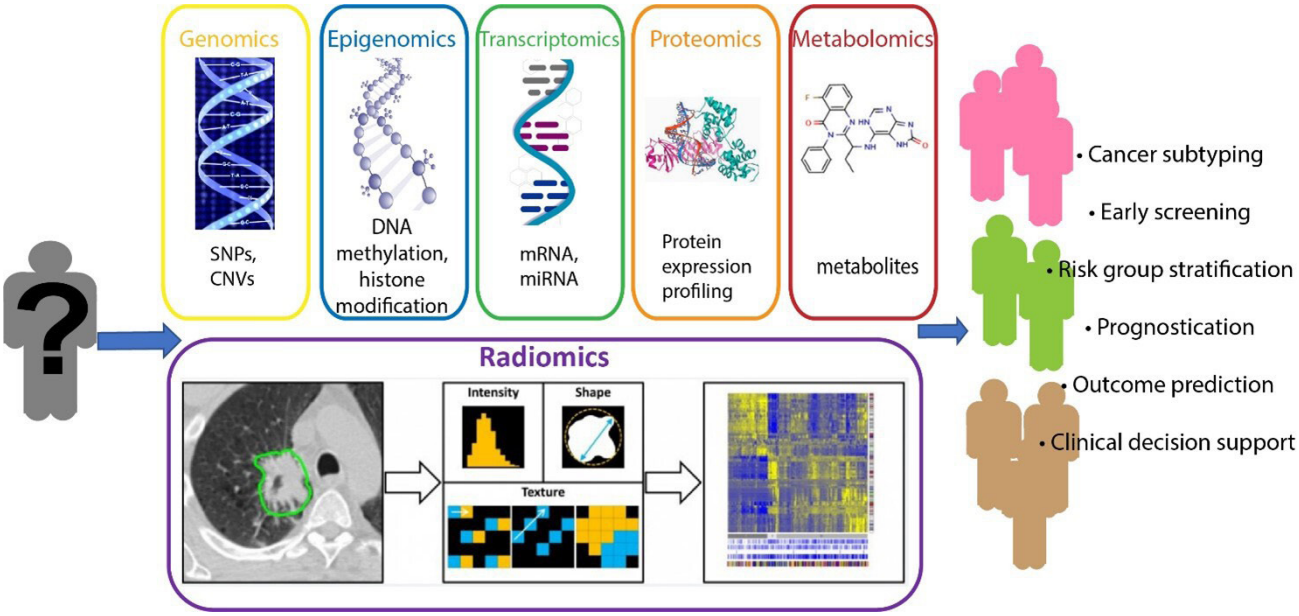
WHAT IS MULTIOMICS?

Omics data include imaging radiomics and various types of biological molecules as shown in Figure 1. Radiomics extracted from medical images has been extensively investigated in cancer research for diagnosis and treatment in recent years. Quantitative imaging features—radiomics play a crucial role in the field of medical research and diagnostics. These features, when evaluated as indicators of biological, physiological, or pathological processes can be considered as imaging biomarkers. Similarly, any biological molecules that are measurable indicators of a sign of a normal or abnormal process, or of a condition or disease can be identified as molecular biomarkers. Molecular biomarkers, which are the backbone of studying biological mechanisms of cancer have been increasingly available due to the recent advances in high throughput omics biotechnologies such as next-generation sequencing, microarray, RNA-seq and mass spectrometry, and are expected to continuously contribute to precision oncology.

Radiomics

Radiomics focuses on relating extracted imaging information to clinical and biological end points in quantitative image analysis.[1] The assumption is that there is useful information embedded in medical images including diagnostic or histopathological images can help with disease diagnosis, prognosis, or prediction. Radiomics has the advantage of being non-invasive compared with traditional

Figure 1. Multiomics data and its role in precision oncology. CNV, copy number variation; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; SNP, single nucleotide polymorphism.



biopsy-based analysis. Recently, the development of quantitative imaging and machine learning (ML) algorithms has moved this field to be integrated with the application of precsion oncology. Conventionally, radiomics features provide information about tumor size, shape, intensity, and texture based on different modalities (CT, MR, PET, molecular imaging etc) and are computed manually. ML methods can then be applied to these features and models can be built for specific end points. It is important to keep the data processing and analysis process consistent to translate these radiomics models as robust tools to the clinic. The image biomarker standardization initiative (IBSI) provides the guidelines for definition, reporting and feature calculation methods.[2] Deep learning (DL) methods avoid the manual feature calculation step and automatically learn features from the images, which becomes more favorable in the radiomics world. Radiomics has been applied to many diseases and different tasks, including cancer detection,[3] prediction of tumor stage,[4] tumor genotypes,[5] tumor response,[6,7] toxicity,[8] local control/failure,[9] and survival[10,11] in oncology.

Molecular biomarkers

Molecular biomarkers include genomics, epigenomics, transcriptomics, proteomics, metabolomics, etc and are collected from various biospecimen such as blood, tissue, tumor, urine, semen, saliva, etc. Genomics refers to the collective characterization and quantification of an organism's genes. The development of high-throughput sequencing technologies which parallelized the sequencing process has greatly increased the availability of genomics data for cancer research. Genomics data[12] is generated from the sequencing of DNA and is commonly used in cancer studies after the necessary annotation and analysis is done. Some examples of genomics data include single nucleotide polymorphisms (SNPs) and copy number variations (CNVs). Epigenomics[13] is a study focusing on the characterization of reversible modification of DNA or histones that affect gene expression without alternating DNA. Epigenetics data include DNA methylation and histone modification. Transcriptomics[14] is a study of the complete set of RNA transcripts. Transcriptome can be obtained by microarray and RNA-seq methods. Some examples of transcriptomics data include protein-coding mRNA, non-coding miRNA, lncRNA, and snRNA. Proteomics[15] focuses on the step after genomics and transcriptomics in the biological system and is a study of proteins. Microarray and mass spectrometry are among the popular methods to gain proteomics data. Metabolomics studies a collective set of small molecules metabolites such as metabolic intermediates, hormones and other signaling molecules to be found in blood and urine biospecimens. Although the metabolome has not been extensively studied for cancer compared to other types of omics data, it also has the potential to be associated with risk in cancer etiology, survival, and overall mortality.[16]

Correlations among different biomarkers

The different omics data may be inherently interconnected in the underlying biological processes, particularly when it comes to molecular biomarkers. Ongoing research efforts also try to link Radiomics with other omics data.

Omics data may correlate in isolation with a specific problem in diagnosis or therapy. However, such an approach may not adequately consider the interactions between different molecular levels.[17] Research has been conducted to reveal interactions and correlations within and across different omics, this may help create cost-effective or less invasive ways to get patient-specific information,[18] discover the new biomarker,[19] uncover biological pathways or mechanisms of cancer,[20] and eventually help provide better diagnosis and therapy of cancer.

ROLES OF MULTIOMICS IN DIAGNOSIS AND TREATMENT OF CANCER

The general goal of omic-type analyses is to improve clinician knowledge about a patient's disease and ultimately contributes to clinical decision-support. Depending on the timing and presentation of this knowledge, the specific application may vary. Initial applications are related to diagnosing a disease and include tumor detection, screening[21] or subtyping.[22] Distinct tumor subtypes, especially those defined by common genetic pathways, can aid risk group stratification[23] and may eventually guide a treatment with targeted agents in novel immunotherapy.[24] Multiomics data along with advanced prognostication,[25–29] and outcome prediction[30,31] can offer possibilities for treatment selection or planning. After treatment has commenced, omics data can be leveraged to monitor response and provide decision-support[32–34] to therapy and adapt if needed.[35] Some examples of modeling designs include: (i) radiomics as predictors for molecular data, and (ii) omics or multiomics data as predictors for a specific clinical end point.

Radiomics, high-throughput quantitative analysis of features derived from medical images, is a logical tool for approaching diagnostic challenges. Imaging is generally available more quickly, easily, and non-invasively than tumor tissue samples needed for genomic analysis. Radiomics has been widely used to distinguish malignant[36] from begin tumors or predict the overall malignancy of visible lesions.[4] However, its wide application is still limited to research, the implementation of radiomics in real clinical settings is in its early stages and requires ongoing and concerted efforts to progress. A focus of radiomic signatures is to predict an already-established genetic or histologic subtype. The rationale is to combine the efficiency and availability of imaging data with the biological implications of established subtypes. Combination of multiple omics data can help uncover better prognostic signatures. For example, in lung cancer genomics, and PET radiomics have been used to predict cancer histotypes and jointly estimate patient prognosis.[37,38]

Other omics analyses for diagnosis include genetic analysis of so-called liquid biopsies from serum samples which are much more accessible than surgical tumor specimens.[39–41] In these samples, genomic, proteomic, and metabolomic analyses are possible on both patient germline DNA as well as circulating tumor DNA. Multiomics approaches are well suited for analyzing liquid biopsy because the samples contain DNA, proteins, and metabolites which can all be mined for useful data.

Multiomics analysis can be used for improving existing cancer therapy and exploring novel therapy. First, the effectiveness of existing treatments can be statistically improved by identifying patients which are unlikely to respond and using alternatives instead. Omics signatures can be used to identify high-risk groups that may be at increased risk for recurrence,[37] and therefore may benefit from treatment escalation. Second, identifying signatures of distinct tumor subtypes allows the development of targeted agents.[42,43] A tremendous amount of effort has been placed on developing differentiated treatment regimens. Roles for multiomics include identifying specific signaling pathways that can be blocked[44] or antigens targetable by immunologic agents.[45] Sometimes, more distinctive prognostic groups can be found specifically with a multiomics approach.[46–48]

In addition to identifying patients who are likely to respond to a treatment beforehand, genomic or radiomic data can monitor ongoing response to treatment. Historically, treatment adaptation has been based on simple metrics like tumor size or visible disease burden. However, complex changes in tumor image texture, proteome, or genome may be useful indicators of response assessment, even just shortly after treatment induction.[49,50] Response assessment is often measured via estimation of local control probability or increased risk of toxicity.[31,50] Examples of adaptation are changes made in chemotherapy regimen or radiation dose escalation.

METHODOLOGY AND VALIDATION

AI/ML/DL methods

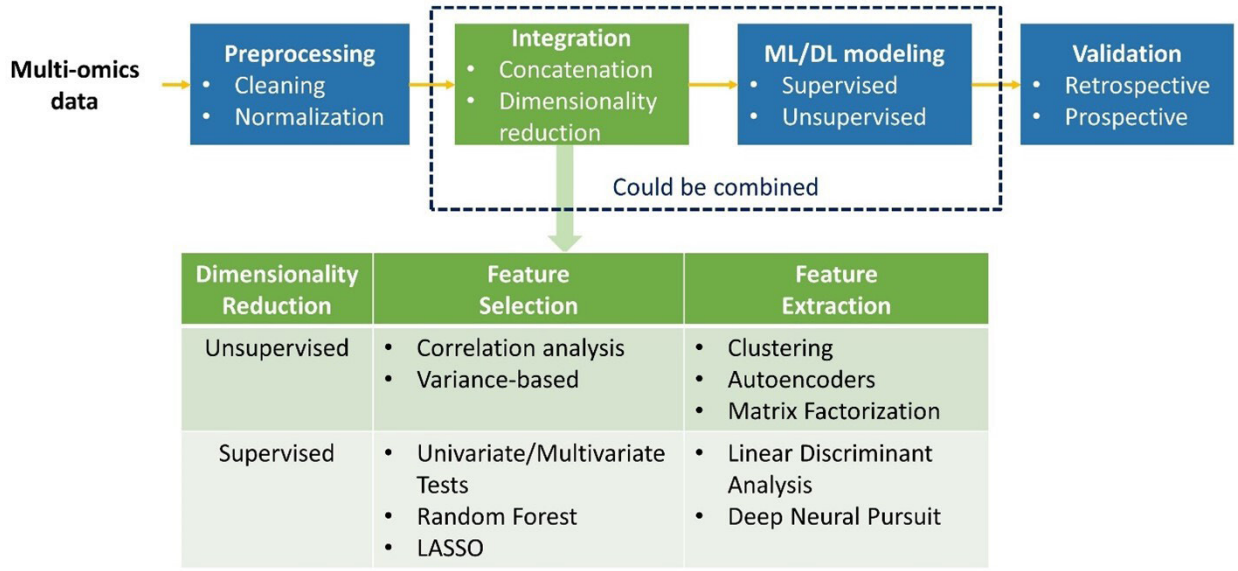
Data integration methods

Integration of high-dimensional and heterogenous multiomics features is a very challenging task. Pre-processing steps such as data cleaning and normalization are essential to ensure data quality and consistency before data integration. Data cleaning normally involves handling missing data using imputation[51] and removing outliers.[52] Popular normalization methods are Z-score normalization and Min-Max scaling. Concatenation of multiomics features is the easiest way to conduct multiomics data integration. However, it may lead to the "curse of dimensionality" problem, *i.e.* the number of features is much larger than the data sample size, and hence model overfitting. It is common to conduct dimensionality reduction before integrating multiomics data for dealing with high-dimensional data.

As shown in Figure 2, dimensionality reduction can be achieved using either feature selection where a subset of original features is selected[53] or feature extraction[54] where representation features are built based on original features. Dimensionality reduction techniques can be further separated into two main categories, supervised or unsupervised, depending on whether data labels are used for feature reduction. Unsupervised feature selection methods can be correlation analysis and variance-based approaches. Supervised feature selection methods use data labels to identify the most relevant features via univariate or multivariate tests, random forest (RF), least absolute shrinkage and selection operator (LASSO) etc. Unsupervised feature extraction methods include clustering,[55] autoencoders,[56] and matrix factorization techniques such as principal component analysis (PCA), and singular value decomposition (SVD). Supervised feature extraction methods include linear discriminant analysis (LDA),[57] deep neural pursuit (DNP)[58] and others. It should be emphasized that validation or testing data sets should be held out from the supervised feature reduction process. Otherwise, it may lead to biased/overpromised model performance.

A combination of different dimensionality reduction methods can also be used to further reduce feature dimension. For example, Sammut et al applied both supervised univariate feature selection and unsupervised collinearity reduction to select features in

Figure 2. Workflow and techniques for multiomics study. DL, deep learning; LASSO, least absolute shrinkage and selection operator; ML, machine learning.



each omic category (genomics, transcriptomics, pathology, and clinical features).[59] The remaining features were concatenated and used to train ML models for predicting neoadjuvant chemotherapy treatment response in breast cancer patients.

ML/DL modeling

ML and DL methods[60] have achieved promising performance in analyzing multiomics data.[9] ML methods build models based on the observed data to make predictions or decisions without being explicitly programmed. DL methods are a subset of ML methods. They are mainly based on artificial neural networks.[61] Conventional ML methods require the input of human-designed features, while DL methods can learn representation directly from raw data, *e.g.* 2D/3D medical images.

ML/DL methods used for multiomics data analysis can also be divided into two categories, unsupervised learning and supervised learning. For example, Zhang et al proposed an unsupervised DL-based clustering method to identify cancer subtypes.[62] Vahabi and Michailidis summarized the application of unsupervised methods in multiomics data analysis to investigate many tasks including disease onset prediction and pathway analysis.[63] On the other hand, supervised learning requires labels for model training to achieve classification or regression tasks such as patient risk prediction, treatment outcome prediction, and survival prediction. Supervised conventional ML methods include generalized linear model (GLM),[64] support vector machine (SVM),[65] and RF. Supervised DL methods could be constructed using multilayer perceptron (MLP),[66] convolutional neural network (CNN),[67] and recurrent neural network (RNN).[68]

It may not be necessary to reduce feature dimensions before integrating multiomic features when using DL methods as they can learn representation from raw features. For example, Cui et al proposed an actuarial DL neural network to predict the probability of radiation-induced pneumonitis and local control in Stage III non-small cell lung cancer patients based on the input of raw multiomics data (dosimetrics, radiomics, proteomics, and transcriptomics).[50] In the proposed model, a 1-D CNN was used to learn representation from dosimetrics, and three variational autoencoders (VAEs)[56] were used to encode the remaining three feature categories, respectively. Encoded features were then fed into a survival neural network (Surv-Net)[69] which was trained in an end-to-end scheme along with the 1D CNN and VAEs.

Validation of models

The risk of bias and potential usefulness of an AI/ML model can be adequately assessed by full and clear reporting of information on all aspects of the model. Multiple guidelines have been published to aid the assessment of AI/ML models. Mongan et al proposed a CLAIM guide[70] to promote clear, transparent and reproducible scientific communication about AI's application in medical images. El Naqa et al[71] proposed a checklist to enhance the reproducibility and generalization of applications of artificial intelligence (AI) in medical physics. Collins et al proposed TRIPOD Initiative[72] as a set of recommendations for the reporting of developing, validating, or updating a prediction model for diagnostic or prognostic purposes.

Retrospective

A retrospective study collects information from patients who have been treated in the past and analyzes the data. Cross-validation (CV) is a popular resampling procedure in retrospective studies used to evaluate ML models and access how the model will perform for an independent test data set. K-fold cross-validation is the most used technique. After equally partitioning the original data set into k subparts or folds, one group is held out for validation and the remaining (k-1) groups are selected as training data. The process is repeated k times until each group has been treated as validation data once. Leave-one-out cross-validation (LOOCV) and stratified k-fold cross-validation are

among the popular variations of CV. As for time-series-related data, time-series CV was proposed to split data into train and validation according to the time. It is also referred to as the forward chaining method or rolling CV, since for a particular iteration, the next instance of train data can be treated as validation data.[73] It is worth noting that when tuning the hyperparameters of the models, it is important to conduct nested CV, otherwise, the intended validation data set will be mistakenly used during the development of the model and thus give an unfair evaluation of the model performance.

Prospective

A prospective study involves following and observing patients over a period of time, collecting patients' information[9] and testing an assumption that was made beforehand. Unlike a retrospective study, a prospective study sets patient eligibility criteria, end points of interest, and follow-up analysis in advance to control for bias and variability. Prospective validation may utilize data from prospective studies and test a pre-existing model. It is highly preferable compared to retrospective validation as hidden confounding factors which may not present or cannot be simulated in the retrospective analysis can eventually emerge in the real world and degrade the performance of the model. Especially, considering the complexity involved in collecting, processing, and analyzing multiomics, prospective validation may give a more straightforward and unbiased evaluation of the findings compared to its retrospective counterpart.

Sometimes, to acquire enough patients in a prolonged follow-up, prospective analysis can be preferably done through collaboration. Gathering data from different institutions may not only accelerate data collection, but also affirm that the research findings can be extended to benefit a larger population, as patients from the same institution may not be representative due to the homogenous clinical practice as well as similar patients' demographics and characteristics. Overall, it is preferable that the validation of the research findings can be conducted in a prospective manner and on an external data set.

APPLICATIONS OF MULTIOMICS

Precision oncology has significantly benefited from the applications of multiomics, encompassing essential areas such as cancer subtyping, genetic mutation prediction, risk group stratification, prognostication, outcome prediction, and clinical decision support. Table 1 provides a comprehensive compilation of exemplary studies illustrating the utilization and impact of multiomics in these crucial aspects of cancer research and treatment.

Cancer subtyping

Cancer is primarily classified by its site of origin. However, cancer occurring at the same site can be genetically different. Thus, subtyping cancers based on their distinct molecular features is necessary for appropriate and effective treatment. Traditionally, pathologists carry out cancer subtyping based on histological data and growth sites. More recently, multiomics data from techniques such as microarray and RNA sequencing (RNA-seq) in conjunction with ML are being used for subtyping.[78] Trivizakis et al[74] applied supervised learning on radiomics, and transcriptomics data obtained from NSCLC patients with well-defined histology subtypes and molecular end points (EGFR and KRAS mutation status). They trained and compared SVM, decision tree, and DNN model for subtypes and mutation prediction, and identified important radiogenomic predictors. Park et al[75] trained and evaluated the performance of various ML models in predicting prognostic biomarkers and molecular subtypes of breast cancer. 18 perfusion parameters of low-dose perfusion CT were analyzed using 5 Ml models to predict lymph node status, tumor grade, tumor size, hormone receptors, HER2, Ki67, and the molecular subtypes. Aftimos et al[76] studied the process of relapse in metastatic breast cancer where they applied an unsupervised uniform manifold approximation and projection (UMAP) clustering algorithm on RNA-seq data to compare gene expression profiles of different cancer subtypes.

Genetic mutation prediction and risk group stratification

Knowledge of the important multiomics features associated with a cancer type and its subtype can be used for genetic mutation prediction and risk group stratification. Shiri et al[21] applied multiple feature selection methods and ML models on radiomics data from NSCLC patients for multivariate prediction of gene mutation status of KRAS and EGFR in NSCLC patients. Since the mutation status infers about NSCLC, such radiomics models can be used for indirect screening. Zhang et al[23] developed ML strategies for risk stratification that identified four molecular subgroups in pediatric medulloblastoma (MB). They extracted 1800 radiomics features from pre-operative MRI images and trained a two-stage classifier which identified non-wingless (WNT) and non-sonic hedgehog (SHH) MB in the first stage and differentiated therapeutically relevant WNT and SHH in the second stage. WNT and SHH are signaling pathways. Additional classifiers for distinguishing Group 3 and Group 4 MB were developed and the identification of radiomics features unique to infantile *vs* childhood SHH subgroups was carried out.

Prognostication of cancer

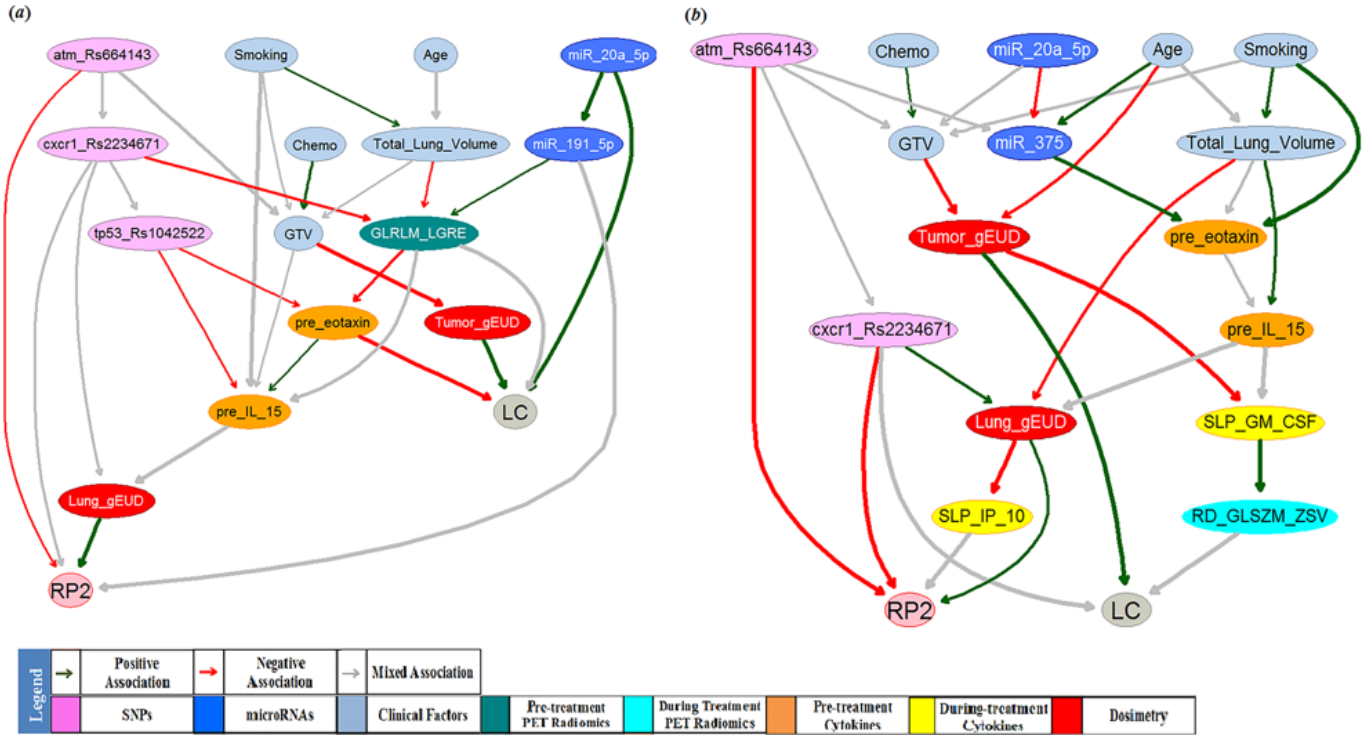
Prognostication provides an estimate of a patient's life expectancy. Besides tailoring treatments, the knowledge of survival time is important for the patient and family in planning their next course of action. Hoivik et al[26] integrated pre-operative MRI with histologic, transcriptomic, and molecular biomarkers to identify aggressive tumor features in endometrial cancer via a survival analysis. ML techniques such as tumor auto segmentation via 3D UNet and K-means unsupervised clustering were applied for the study. Zeng et al[27] used CT radiomics features for predicting genetic mutations and mRNA-based molecular subtypes and then combined radiomic features with molecular data for 5-year overall survival prediction in clear cell renal cell carcinoma. They analyzed four feature selection methods and eight classifier models, out of which RF classifier performed the best regardless of the feature selection method. They found out that integrative multiomics (radiomics, genomics, transcriptomics, and proteomics) data have the best OS prediction power followed by radiomics + omics, and single omics. OS analysis via multiomics data and advanced ML modeling is rapidly

Table 1. Summary of different applications of multiomics analysis

Application	Study	Objective	Disease	Omics data	ML/DL model
Cancer Subtyping	2021, Trivizakis et al.[74]	Assessing molecular and histology subtypes	NSCLC	Radiomics, Transcriptomics (CT, PET/CT, RNA-seq, genomic, histology)	SVM, DT, DNN
	2019, Park et al.[75]	Predict prognostic biomarkers and molecular subtypes	Breast cancer	Radiomics: Perfusion CT and histopathological reports	SVM, DT including RF, NB, DNN
	2021, Aftimos et al.[76]	Study the processes of relapse in metastatic breast cancer (*Note: Cancer subtyping one of the goals*)	Breast cancer	Genomic and Transcriptomic (RNA-seq, targeted gene seq, SNP array)	UMAP
Genetic mutation prediction and risk group stratification	2020, Shiri et al.[21]	Predict EGFR and KRAS mutation status for early screening	NSCLC	Radiomics and Genomic from TCIA (PET, low-dose CT, CTD, EGFR, KRAS)	GLM, NB, k-NN, DT including ensemble methods, SVM
	2022, Zhang et al.[23]	Identify four clinically significant MB molecular subgroups	Pediatric medulloblastoma	Radiomics and Molecular Data (T_2- and T_1 weighted pre-operative MRI, fluorescence, methylation, next-generation sequencing)	SVM, LR, k-NN, RF, XGBoost, NN
Prognostication of cancer	2021, Hoivik et al.[26]	Identify aggressive tumor features in endometrial cancer patients for accurate diagnosis and tailored treatment	Endometrial cancer	Radiomics and Genomics (pre-operative MRI, histologic-, transcriptomic- and molecular biomarkers)	3D UNet (CNN), K-means Clustering
	2021, Zeng et al.[27]	Feasibility study of molecular characteristics prediction and OS prediction	ccRCC	Radiomics, Genomics, Transcriptomics, and Proteomics (genetic data from TCGA and contrast-enhanced CT from TCIA)	GBDT, LASSO, RF, XGBoost GBDT, AdaBoost, LR, DT, SVM, NB, K-NN
	2022, Meneghetti et al.[28]	Subtype classification and improved outcome prognosis of patients	HNSCC	Radiomics and Whole transcriptome (CT, microarray)	LR, GSVA, PCA
	2022, Iwatate et al.[29]	Predict genetic information (p53 status, PD-L1 expression), and prognosis	Pancreatic cancer	Radiomics and proteomics (CT scan, MRI & PET when available, immunohistochemistry of p53 and PD-L1)	RF, XGBoost
	2021, Poirion et al.[25]	Efficient data integration of mutliomics data for prognosis and survival prediction	Liver and breast cancer	Genomics (RNA-Seq, miRNA Sequencing, DNA methylation from TCGA)	Autoencoder (DNN), GM, SVM
Outcome Prediction	2022, Luo et al.[77]	Accurate RT outcome prediction via interpretable and explainable models with human-in-the-loop	NSCLC, HCC	Radiomics and Genomics, (CT, PET, cytokines, miRNA, SNP, dosimetry)	MB, BN
	2021, Cui et al.[50]	Joint and independent RT outcome prediction of local control and toxicities	NSCLC	Radiomics and Genomics, (CT, PET, cytokines, miRNA, SNP, DVH)	CNN, VAE, Surv-Net
Clinical Decision Support	2022, Niraula et al.[33]	Develop a decision-support system for AI-assisted decision-making for optimal outcome in response-adaptive RT	NSCLC, HCC	Radiomics and Genomics, (CT, PET, cytokines, miRNA, SNP, dosimetry)	DNN, GNN, DRL, qDRL

AdaBoost, Adaptive Boosting; BN, Bayesian Network; CNN, Convolution NN; CT, Computed Tomography; CTD, Contrast-enhanced diagnostic quality CT; DNN, Deep Neural Network; DRL, Deep Reinforcement Learning; DT, Decision Tree; GBDT, Gradient Boosting Decision Tree; GLM, Generalized Linear Model; GM, Gaussian Mixer; GNN, Graph NN; GSVA, Gene Set Variation Analysis; HCC, Hepatocellular Carcinoma; HNSCC, Head and neck squamous cell carcinoma; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; MB, Markov Blanket; MRI, Magnetic Resonance Imaging; NB, Naïve Bayes; NSCLC, Non-small-cell lung carcinoma; PCA, Principal Component Analysis; PET, Positron Emission Tomography; RF, Random Forest; SNP, Single Nucleotide Polymorphism; SVM, Support Vector Machine; TCGA, The Cancer Genome Atlas; TCIA, The Cancer Imaging Archive; UMAP, Uniform Manifold Approximation and Projection; VAE, Variation Autoencoders; XGBoost, Extreme Gradient Boost; ccRCC, Clear cell renal cell carcinoma; k-NN, k-Nearest Neighbor; qDRL, Quantum DRL.

Figure 3. Pre- (**a**) and during (**b**) treatment situation awareness Bayesian networks for the joint prediction of lung cancer patients' tumor local control and radiation pneumonitis with Grade 2 after radiation therapy based on the discovery data set. miR, microRNA; GLRLM_LGRE, Gray Level Run Length Matrix_low gray level run emphasis; gEUD, generalized equivalent uniform dose; LC, local control; RP2, radiation pneumonitis grade equal or greater than 2; GTV, gross tumor volume; SLP, slope; IL, interleukin; IP, Interferon-gamma-Inducible Protein.



advancing in other cancer types, such as HNSCC,[28] pancreatic cancer,[29] liver and breast cancer,[25] etc.

Outcome prediction

Accurate outcome prediction helps in personalizing treatment and increasing treatment efficacy. Information-rich multiomics data have been shown to improve model performance and help lower decision-making uncertainty.[32] Luo et al[31,79] have employed an interpretable Bayesian Network approach on multiomics data for predicting local control and radiation-induced complications in NSCLC. They have used the Markov blanket approach and pre-existing expert knowledge in feature selection and building a relational graph between the selected features and the outcome as shown in Figure 3. Cui et al[50] developed a novel actuarial DNN for independent and joint prediction of local control and radiation pneumonitis in NSCLC. In their design, differential dose–volume histograms were fed into 1D convolution neural network, and high-dimensional imaging and molecular data were fed into a variational autoencoder. The reduced representation from CNN and Autoencoder was then fed into Surv-Nets for predicting time-to-event probabilities.

Clinical decision-support

An outcome prediction model can be employed to develop a decision-support system that maximizes local control and minimizes toxicities.[77,80] Niraula et al[33,81] have developed decision-making software for adaptive radiotherapy that can take in patients' pre- and mid-treatment multiomics data and recommend optimal dose decisions for the rest of the treatment period. The decision-making tool consists of two main AI components: (i) artificial radiotherapy environment (ARTE) and (ii) optimal decision-maker (ODM). They have used Graphical Neural Networks for designing ARTE and Deep Reinforcement learning for ODM. Additionally, quantum deep reinforcement learning as ODM[82] and the inclusion of Gaussian process in ARTE as uncertainty estimating and decision improving tool[34] have been explored.

TRANSITION INTO THE REAL WORLD

Multiple challenges remain to be resolved in order to translate multiomics studies into real-world clinical applications. Firstly, heterogenous multiomics data may originate from different sources or varied technologies and may contain missing or mislabeled information, these may affect the generalizability and reproducibility of the multiomics study. Second, multiomics data may not be readily available in current routine clinical practice, this demands data management and data sharing of multiomics that requires efforts in building infrastructures for storing, sharing, and re-using omics data. Thirdly, data fusion methods that can integrate multiomics data holistically and systematically are still lacking. Fourthly, clinical applications also require the interpretability of AI/ML methods which may advance the understanding of cancer biology and more importantly, help gain the trust of clinicians and build safeguard mechanisms. Ultimately, multiomics clinical trials are expected to be done to

establish robust connections between multiomics research findings and appropriate clinical decisions.

Standardization of sample collection and analysis pipelines

The success of a multiomics project strongly depends on the cooperation between a carefully designed experimental setup and a suitable data processing pipeline. As multiomics data sets may contain very different data modalities originating from varied technologies, platforms and analyzing tools. Also, missing data and mislabeled data are common problems in large-scale multiomics study.

Tarazona et al proposed a set of harmonized Figures of Merit (FoM) as quality descriptors applicable to different omic data types. Based on the FoM, they formulated the MultiPower method to estimate the optimal sample size in a multiomics experiment and support different experimental settings, data types and sample sizes.[83] The FoM definitions constitute a framework for quality control and precision in the design of multiomic experiments.

Benchmarking computational tools

In multiomics studies, benchmarking study may be used to determine the strengths of computational methods and select suitable methods for multiomics analysis. Mangul et al[84] proposed the following principles for rigorous, reproducible, transparent, and systematic benchmarking. Some recommendations include compiling a comprehensive list of tools, preparing and describing benchmarking data, selecting evaluation metrics, considering parameter optimization, summarizing algorithm features, sharing commands for installing and running tools, defining a universal format (if necessary) and providing a flexible interface for downloading data. In addition to the above principles, Weber et al summarized other key practical guidelines and recommendations for performing high-quality benchmarking analyses, for instance, evaluating secondary measures including computational requirements as well as designing the benchmark to enable future extensions and following reproducible research best practices by making code and data publicly available.[85]

Availability of omics data and data sharing

Currently, omics data are not easily accessible in routine clinical practice. Genomics data are occasionally collected during the cancer diagnosis, but other omics data such as transcriptomics and proteomics data are still behind. This is expected to change in the future, as more multiomics research findings become readily translated into clinical practice for precision oncology.[86]

To enhance the reusability of existing data, Wilkinson et al proposed the principles of findability, accessibility, interoperability, and reusability, named FAIR principles. In addition to supporting data reuse by individuals, the FAIR principles also put specific emphasis on enhancing the ability of machines to automatically find and use the data. GitHub appears to be the most popular platform for FAIR[87] sharing of data and code, and Bioconductor and comprehensive R archive network (CRAN) are also among the popular platforms.

Currently, there are many ongoing efforts to enhance multiomics data sharing in cancer research. The cancer genome atlas (TCGA) is a landmark multiomics data set.[88] It includes genomic, epigenomic, transcriptomic, and proteomic for various types of cancers. International cancer genome consortium (ICGC)[89] is an international collaboration where members from different countries can share omics data (mostly mutation-related genomic alterations) of various tumor types. And, therapeutically applicable research to generate effective treatments (TARGET)[90] which aims to determine molecular changes that drive childhood cancers containing omics data such as gene expression and miRNA.

Efforts have also been made to reveal the connections among omics datasets from different resources. Perez-Riverol et al released the first version of the omics discovery index (OmicsDI) in 2017 as a lightweight system to aggregate datasets across multiple public omics data resources. It aggregates genomics, transcriptomics, proteomics, metabolomics and multiomics data sets, as well as computational models of biological processes.[91] In addition, Zenodo was proposed to allow users to upload raw data files, tables, figures, and code. In addition to providing digital object identifiers (DOIs), it supports code repositories, with GitHub integration. Moreover, Foster et al proposed an open science framework (OSF) to promote open, centralized workflows by enabling the capture of different aspects and products of the research lifecycle, including developing a research idea, designing a study, storing, and analyzing collected data, and writing and publishing reports or paper. The OSF also supports DOIs while promoting open-source sharing that adheres to the FAIR guidelines.[92]

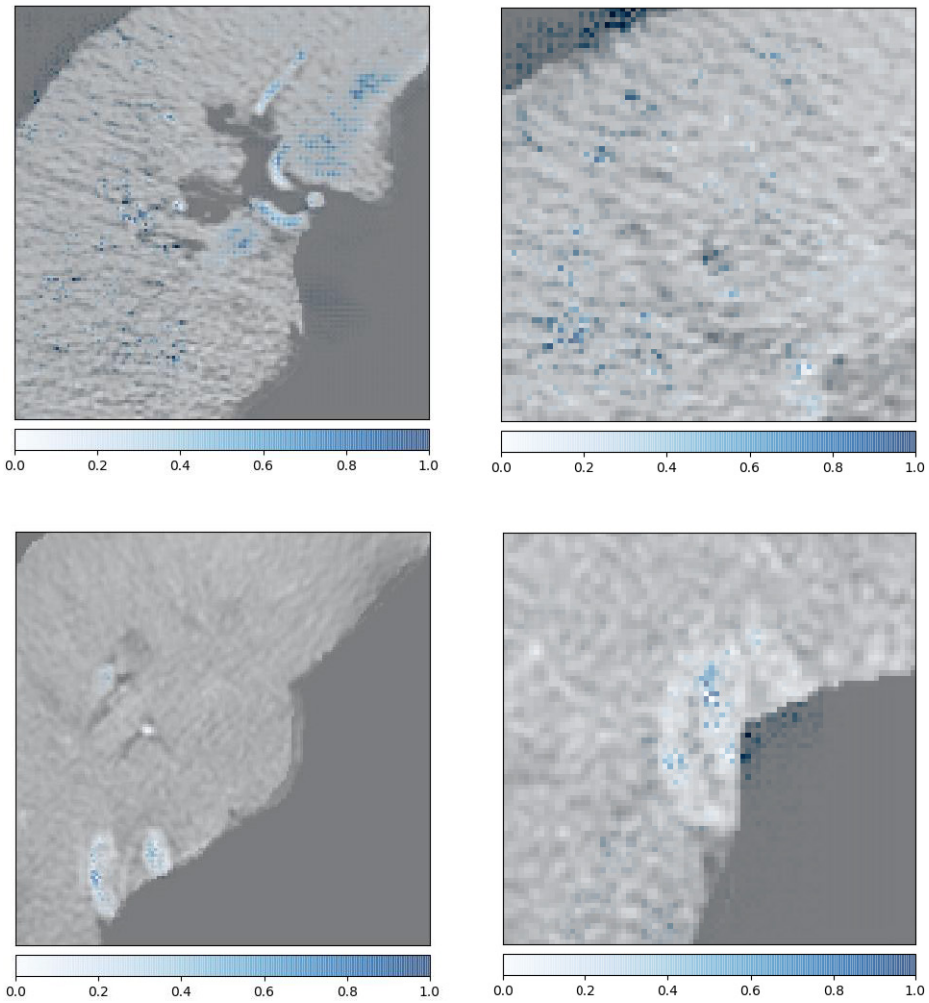
Data fusion

Different categories of omics data can complement each and provide an integrative picture of cancer development and progression. Additionally, one may reduce the experimental or biological uncertainty in the analysis by combining multiomics data. However, integrating omics data from different sources can be challenging, as heterogenous multiomics data may be generated by varied technologies and may have a variety of magnitude. As discussed in the method section, data normalization and scaling should be adapted to pre-process omics-data before integrating. And, data representations and dimensionality reduction methods should be used when combining a large number of omics data. Most omics integration methods do not imply any hierarchy among the data, which may preclude understanding cancer biology in a holistic and systematic manner. Hence, integration of multiomics data from different levels of a biological system using some advanced methods such as Bayesian network[31] and Similarity network fusion[93] can enforce hierarchy and incorporate any prior information and expert knowledge when integrating omics data from different sources.

Interpretability

The development of ML algorithms shows great potential for medical applications, including multiomics analysis. However, one of the biggest challenges is the "black box" nature of these complicated models. It is crucial for clinicians and patients to

Figure 4. Example images for low/high liver-GTV GLSZE–GLN values. Top row shows the images for a patient with GLN 0.756, died in 264 days. Bottom row shows the images of a patient with GLN 0.154, survived 760 days (right censored). The left column shows most of the liver tissues. The right column shows the zoomed in tissues. The blue dots are the output of integrated gradients method that shows the critical pixels for the prediction of the neural network. GLSZE: gray level size zone, GLN: gray level non-uniformity.



understand the models before adopting them in the clinic. Interpretable AI has been given a lot of attention, especially in the medical field[94] to enhance the understanding of the models.[95] In general, there are two types of interpretable methods: inherent interpretable methods, such as decision tree, logistic regression, naïve Bayes, etc; and post-training methods that try to explain the built "black box" models (*e.g.* deep neural networks).[96] For the post-hoc interpretation, ablation analysis is commonly used for DL, which removes certain data points (structures, nodes) to see how this will affect the model accuracy.[97] Gradient-based methods are another type of popular interpretation methods. Layer-wise relevance propagation (ϵ-LRP),[98] DeepLIFT,[99] and Integrated Gradients (IG)[100] methods are some examples of gradient-based methods. These interpretable methods have been used in medical applications widely. Saliency maps on chest radiographs have been used to validate the learned patterns of a DL system classifying patients as having pneumonia.[101] As shown in Figure 4, integrated gradients were used in one study for risk prediction of overall survival (OS) in hepatocellular carcinoma (HCC) patients treated with stereotactic body radiation therapy.[10] Another study predicting 10-year prostate cancer mortality using XGBoost applied Shapley values and provides interpretability to understand the results.[102]

Multi-institutional prospective clinical trials

One of the challenges for the multiomics study is the limited number of samples which may lead to high false-positive rate as well as poor reproducibility and generalizability of the findings. In addition to previously mentioned initiatives such as standardization of data collection and analysis, transparency of reporting, and data sharing, multi-institutional prospective clinical trials may eventually establish robust connections between research findings and appropriate treatment intervention of cancer. Data from a single institution may lack variability in patients' demographics or clinical practices which can potentially introduce severe systematic bias to the multiomics research, hence, validating multiomics research findings with multiple external data sets is highly recommended. This is by no means a trivial

task, there are multiple challenges existing such as the standardization of analysis platforms, building of centralized computer infrastructure, and resolving ethics issues relating to adopting AI techniques into clinical practice.

In addition, to prospectively collect the validation data, the prospective validation of research findings can be done by carefully designing a randomized controlled clinical. In such a randomized controlled clinical trial, patients are assigned to treatment groups with or without innervations based on multiomics. The credibility of the findings can then be verified by comparing pre-determined outcomes in the two groups. Randomized controlled trials will balance participants' characteristics between groups, providing a rigorous and unbiased tool for causal inference, hence it is regarded as the most ideal way to prospectively validate multiomics research findings. Recent years have witnessed some pioneering work[103,104] of prospectively validating the AI models in clinical settings in oncology, we expect this trend continues and more work related to the multiomics study will be evaluated prospectively and eventually translate into real clinics and contribute to precision oncology.

CONCLUSION

Multiomics data have been increasingly investigated for better diagnosis and therapy in the era of precision oncology. AI including ML and DL techniques combined with the exponential growth of multiomics data may have great potential to revolutionize cancer subtyping, risk stratification, prognostication, prediction and clinical decision-making. However, the translation of multiomics research to real clinics still requires consistent efforts in standardizing omics data collection and analysis, building computational infrastructure for data sharing and storing, developing advanced methods to improve data fusion and interpretability, and ultimately, conducting large-scale prospective clinical trials to fill the gap between study findings and clinical benefits.

FUNDING

R01 CA233487.

REFERENCES

1. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. *Med Phys* 2020; **47**: e185–202. https://doi.org/10.1002/mp.13678
2. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; **295**: 328–38. https://doi.org/10.1148/radiol.2020191145
3. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A fully-automatic multiparametric radiomics model: Towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci Rep* 2017; **7**(1). https://doi.org/10.1038/s41598-017-14753-7
4. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5**(1): 4006. https://doi.org/10.1038/ncomms5006
5. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, et al. Deciphering Genomic underpinnings of quantitative MRI-based Radiomic phenotypes of invasive breast carcinoma. *Sci Rep* 2015; **5**(1). https://doi.org/10.1038/srep17787
6. Wei L, Cui C, Xu J, Kaza R, El Naqa I, Dewaraja YK. Tumor response prediction in 90Y radioembolization with PET-based radiomics features and absorbed dose Metrics. *EJNMMI Phys* 2020; **7**(1): 74. https://doi.org/10.1186/s40658-020-00340-9
7. Leijenaar RT, Bogowicz M, Jochems A, Hoebers FJ, Wesseling FW, Huang SH, et al. Development and validation of a radiomic signature to predict HPV (P16) status from standard CT imaging: A multicenter study. *Br J Radiol* 2018; **91**: 20170498. https://doi.org/10.1259/bjr.20170498
8. Cunliffe A, Armato SG III, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *International Journal of Radiation Oncology*Biology*Physics* 2015; **91**: 1048–56. https://doi.org/10.1016/j.ijrobp.2014.11.030
9. Cui S, Hope A, Dilling TJ, Dawson LA, Ten Haken R, El Naqa I. Artificial intelligence for outcome modeling in radiotherapy. *Semin Radiat Oncol* 2022; **32**: 351–64. https://doi.org/10.1016/j.semradonc.2022.06.005
10. Wei L, Owen D, Rosen B, Guo X, Cuneo K, Lawrence TS, et al. A deep survival interpretable radiomics model of hepatocellular carcinoma patients. *Phys Med* 2021; **82**: 295–305. https://doi.org/10.1016/j.ejmp.2021.02.013
11. Oikonomou A, Khalvati F, Tyrrell PN, Haider MA, Tarique U, Jimenez-Juan L, et al. Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. *Sci Rep* 2018; **8**(1): 4003. https://doi.org/10.1038/s41598-018-22357-y
12. Chakravarty D, Solit DB. Clinical cancer genomic profiling. *Nat Rev Genet* 2021; **22**: 483–501. https://doi.org/10.1038/s41576-021-00338-8
13. Jones PA, Issa JPJ, Baylin S. targeting the cancer epigenome for therapy. *Nat Rev Genet* 2016; **17**: 630–41. https://doi.org/10.1038/nrg.2016.93
14. Goodall GJ, Wickramasinghe VO. RNA in cancer. *Nat Rev Cancer* 2021; **21**: 22–36. https://doi.org/10.1038/s41568-020-00306-0
15. Kwon YW, Jo H-S, Bae S, Seo Y, Song P, Song M, et al. Application of proteomics in cancer: Recent trends and approaches for biomarkers discovery. *Front Med (Lausanne)* 2021; **8**: 747333. https://doi.org/10.3389/fmed.2021.747333
16. Loftfield E, Rothwell JA, Sinha R, Keski-Rahkonen P, Robinot N, Albanes D, et al. Prospective investigation of serum metabolites, coffee drinking, liver cancer incidence, and liver disease mortality. *J Natl Cancer Inst* 2020; **112**: 286–94. https://doi.org/10.1093/jnci/djz122
17. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights* 2020; **14**: 1177932219899051. https://doi.org/10.1177/1177932219899051
18. Aoude LG, Wong BZY, Bonazzi VF, Brosda S, Walters SB, Koufariotis LT, et al.

Radiomics biomarkers correlate with Cd8 expression and predict immune signatures in melanoma patients. *Mol Cancer Res* 2021; **19**: 950–56. https://doi.org/10.1158/1541-7786.MCR-20-1038
19. Chen X, Li H, Mandic M, Hoffmeister M, Brenner H. Assessment of body mass index, polygenic risk score, and development of colorectal cancer. *JAMA Netw Open* 2022; **5**: e2248447. https://doi.org/10.1001/jamanetworkopen.2022.48447
20. Dashti H, Dehzangi I, Bayati M, Breen J, Beheshti A, Lovell N, et al. Integrative analysis of Mutated genes and mutational processes reveals novel mutational biomarkers in colorectal cancer. *BMC Bioinformatics* 2022; **23**(1): 138. https://doi.org/10.1186/s12859-022-04652-8
21. Shiri I, Maleki H, Hajianfar G, Abdollahi H, Ashrafinia S, Hatt M, et al. Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. *Mol Imaging Biol* 2020; **22**: 1132–48. https://doi.org/10.1007/s11307-020-01487-8
22. Menyhárt O, Győrffy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput Struct Biotechnol J* 2021; **19**: 949–60. https://doi.org/10.1016/j.csbj.2021.01.009
23. Zhang M, Wong SW, Wright JN, Wagner MW, Toescu S, Han M, et al. MRI radiogenomics of pediatric medulloblastoma: A multicenter study. *Radiology* 2022; **304**: 406–16. https://doi.org/10.1148/radiol.212137
24. Wu J, Mayer AT, Li R. Integrated imaging and molecular analysis to decipher tumor microenvironment in the era of Immunotherapy. *Semin Cancer Biol* 2022; **84**: 310–28. https://doi.org/10.1016/j.semcancer.2020.12.005
25. Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. Deepprog: An ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med* 2021; **13**(1): 112. https://doi.org/10.1186/s13073-021-00930-x
26. Hoivik EA, Hodneland E, Dybvik JA, Wagner-Larsen KS, Fasmer KE, Berg HF, et al. A radiogenomics application for prognostic profiling of endometrial cancer. *Commun Biol* 2021; **4**(1): 1363. https://doi.org/10.1038/s42003-021-02894-5
27. Zeng H, Chen L, Wang M, Luo Y, Huang Y, Ma X. Integrative radiogenomics analysis for predicting molecular features and survival in clear cell renal cell carcinoma. *Aging* 2021; **13**: 9960–75. https://doi.org/10.18632/aging.202752
28. Rabasco Meneghetti A, Zwanenburg A, Linge A, Lohaus F, Grosser M, Baretton GB, et al. Integrated radiogenomics analyses allow for subtype classification and improved outcome prognosis of patients with locally advanced HNSCC. *Sci Rep* 2022; **12**(1): 16755. https://doi.org/10.1038/s41598-022-21159-7
29. Iwatate Y, Hoshino I, Yokota H, Ishige F, Itami M, Mori Y, et al. Radiogenomics for predicting P53 status, PD-L1 expression, and prognosis with machine learning in pancreatic cancer. *Br J Cancer* 2020; **123**: 1253–61. https://doi.org/10.1038/s41416-020-0997-1
30. Cui S, Luo Y, Hsin Tseng H, Ten Haken RK, El Naqa I. Artificial neural network with composite architectures for prediction of local control in radiotherapy. *IEEE Trans Radiat Plasma Med Sci* 2019; **3**: 242–49. https://doi.org/10.1109/TRPMS.2018.2884134
31. Luo Y, Jolly S, Palma D, Lawrence TS, Tseng H-H, Valdes G, et al. A situational awareness bayesian network approach for accurate and credible personalized adaptive radiotherapy outcomes prediction in lung cancer patients. *Physica Medica* 2021; **87**: 11–23. https://doi.org/10.1016/j.ejmp.2021.05.032
32. Niraula D, Cui S, Pakela J, Wei L, Luo Y, Ten Haken RK, et al. Current status and future developments in predicting outcomes in radiation oncology. *Br J Radiol* 2022; **95**: 20220239. https://doi.org/10.1259/bjr.20220239
33. Niraula D, Sun W, Jin J, Dinov ID, Cuneo K, Jamaluddin J, et al. n.d.).(A clinical decision support system for AI-assisted decision-making in response-adaptive radiotherapy (arclids). *Sci Rep*; **13**(1). https://doi.org/10.1038/s41598-023-32032-6
34. Sun W, Niraula D, El Naqa I, Ten Haken RK, Dinov ID, Cuneo K, et al. Precision radiotherapy via information integration of expert human knowledge and AI recommendation to optimize clinical decision making. *Computer Methods and Programs in Biomedicine* 2022; **221**: 106927. https://doi.org/10.1016/j.cmpb.2022.106927
35. Cui S, El Naqa I. n.d.).(Prediction of oncology treatment outcomes. In: *Machine and Deep Learning in Oncology, Medical Physics and Radiology*. 2nd ed. Springer International Publishing
36. Verma V, Simone CB II, Krishnan S, Lin SH, Yang J, Hahn SM. The rise of radiomics and implications for oncologic management. *J Natl Cancer Inst* 2017; **109**(7). https://doi.org/10.1093/jnci/djx055
37. Kirienko M, Sollini M, Corbetta M, Voulaz E, Gozzi N, Interlenghi M, et al. Radiomics and gene expression profile to characterise the disease and predict outcome in patients with lung cancer. *Eur J Nucl Med Mol Imaging* 2021; **48**: 3643–55. https://doi.org/10.1007/s00259-021-05371-7
38. Nair VS, Gevaert O, Davidzon G, Napel S, Graves EE, Hoang CD, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res* 2012; **72**: 3725–34. https://doi.org/10.1158/0008-5472.CAN-11-3943
39. Wang Y, Zhang C, Zhang P, Guo G, Jiang T, Zhao X, et al. Serum exosomal microRNAs combined with alpha-fetoprotein as diagnostic markers of hepatocellular carcinoma. *Cancer Med* 2018; **7**: 1670–79. https://doi.org/10.1002/cam4.1390
40. Cohen JD, Javed AA, Thoburn C, Wong F, Tie J, Gibbs P, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc Natl Acad Sci U S A* 2017; **114**: 10202–7. https://doi.org/10.1073/pnas.1704961114
41. Tie J, Kinde I, Wang Y, Wong HL, Roebert J, Christie M, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann Oncol* 2015; **26**: 1715–22. https://doi.org/10.1093/annonc/mdv177
42. Tixier F, Cheze-le-Rest C, Schick U, Simon B, Dufour X, Key S, et al. Transcriptomics in cancer revealed by positron emission tomography radiomics. *Sci Rep* 2020; **10**(1): 5660. https://doi.org/10.1038/s41598-020-62414-z
43. Heo YJ, Hwa C, Lee GH, Park JM, An JY. Integrative multi-omics approaches in cancer research: from biological networks to clinical subtypes. *Mol Cells* 2021; **44**: 433–43. https://doi.org/10.14348/molcells.2021.0042
44. Dekker LJM, Kannegieter NM, Haerkens F, Toth E, Kros JM, Steenhoff Hov DA, et al. Multiomics profiling of paired primary and recurrent glioblastoma patient tissues. *Neurooncol Adv* 2020; **2**(1): vdaa083. https://doi.org/10.1093/noajnl/vdaa083
45. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 2019; **177**: 1035–1049.. https://doi.org/10.1016/j.cell.2019.03.030
46. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes

of colorectal cancer. *Nat Med* 2015; **21**: 1350–56. https://doi.org/10.1038/nm.3967
47. Li L, Wei Y, To C, Zhu C-Q, Tong J, Pham N-A, et al. Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. *Nat Commun* 2014; **5**(1). https://doi.org/10.1038/ncomms6469
48. Michaut M, Chin S-F, Majewski I, Severson TM, Bismeijer T, de Koning L, et al. Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* 2016; **6**: 18517. https://doi.org/10.1038/srep18517
49. Satpathy S, Jaehnig EJ, Krug K, Kim B-J, Saltzman AB, Chan DW, et al. Microscaled Proteogenomic methods for precision oncology. *Nat Commun* 2020; **11**(1): 532. https://doi.org/10.1038/s41467-020-14381-2
50. Cui S, Ten Haken RK, El Naqa I. Integrating multiomics information in deep learning architectures for joint actuarial outcome prediction in non-small cell lung cancer patients after radiation therapy. *Int J Radiat Oncol Biol Phys* 2021; **110**: 893–904. https://doi.org/10.1016/j.ijrobp.2021.01.042
51. Liao SG, Lin Y, Kang DD, Chandra D, Bon J, Kaminski N, et al. Missing value imputation in high-dimensional Phenomic data: imputable or not, and how *BMC Bioinformatics* 2014; **15**(1): 346. https://doi.org/10.1186/s12859-014-0346-6
52. Domingues R, Filippone M, Michiardi P, Zouaoui J. A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition* 2018; **74**: 406–21. https://doi.org/10.1016/j.patcog.2017.09.037
53. Rancati T, Fiorino C. Modelling radiotherapy side effects. In: *Building a Predictive Model of Toxicity: Methods*. Boca Raton, FL : CRC Press, Taylor & Francis Group, [2019] |: CRC Press, Taylor&Francis Group; 2019. https://doi.org/10.1201/b21956
54. Cui S, Luo Y, Tseng HH, Ten Haken RK, El Naqa I. Combining handcrafted features with latent variables in machine learning for prediction of radiation-induced lung damage. *Med Phys* 2019; **46**: 2497–2511. https://doi.org/10.1002/mp.13497
55. Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. *Neurocomputing* 2017; **267**: 664–81. https://doi.org/10.1016/j.neucom.2017.06.053
56. Kigma D, Velling M. Auto-Encoding Variational Bayes. In: ; 2013.
57. Vogelstein JT, Bridgeford EW, Tang M, Zheng D, Douville C, Burns R, et al. Supervised Dimensionality reduction for big data. *Nat Commun* 2021; **12**(1): 2872. https://doi.org/10.1038/s41467-021-23102-2
58. Liu B, Wei Y, Zhang Y, Yang Q. Deep neural networks for high dimension, low sample size data. In: Paper presented at the In: Paper presented at the In: Paper presented at the In: Paper presented at the Twenty-Sixth International Joint Conference on Artificial Intelligence, , California, Melbourne, Australia. https://doi.org/10.24963/ijcai.2017/318
59. Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* 2022; **601**: 623–29. https://doi.org/10.1038/s41586-021-04278-5
60. Cui S, Tseng HH, Pakela J, Ten Haken RK, El Naqa I. Introduction to machine and deep learning for medical physicists. *Med Phys* 2020; **47**: e127–47. https://doi.org/10.1002/mp.14140
61. Jain AKJianchangMMohiuddin KM. Artificial neural networks: A tutorial. *Computer* 1996; **29**: 31–44. https://doi.org/10.1109/2.485891
62. Zhang Y, Kiryu H. MODEC: An unsupervised clustering method integrating omics data for identifying cancer subtypes. *Brief Bioinformatics* 2022; **23**(6). https://doi.org/10.1093/bib/bbac372
63. Vahabi N, Michailidis G. Unsupervised multi-Omics data integration methods: A comprehensive review. *Front Genet* 2022; **13**: 854752. https://doi.org/10.3389/fgene.2022.854752
64. McCullagh P, Nelder JA. Generalized linear models. In: *Generalized Linear Models*. ; 2019. https://doi.org/10.1201/9780203753736
65. Noble WS. What is a support vector machine *Nat Biotechnol* 2006; **24**: 1565–67. https://doi.org/10.1038/nbt1206-1565
66. Murtagh F. Multilayer Perceptrons for classification and regression. *Neurocomputing* 1991; **2**: 183–97. https://doi.org/10.1016/0925-2312(91)90023-5
67. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018; **9**: 611–29. https://doi.org/10.1007/s13244-018-0639-9
68. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena* 2020; **404**: 132306. https://doi.org/10.1016/j.physd.2019.132306
69. Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks. *PeerJ* 2019; **7**(1): e6257. https://doi.org/10.7717/peerj.6257
70. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiol Artif Intell* 2020; **2**(2): e200029. https://doi.org/10.1148/ryai.2020200029
71. El Naqa I, Boone JM, Benedict SH, Goodsitt MM, Chan H, Drukker K, et al. AI in medical physics: guidelines for publication. *Med Phys* 2021; **48**: 4711–14. https://doi.org/10.1002/mp.15170
72. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162**: 735–36. https://doi.org/10.7326/L15-5093-2
73. Bergmeir C, Benítez JM. On the use of cross-validation for time series predictor evaluation. *Information Sciences* 2012; **191**: 192–213. https://doi.org/10.1016/j.ins.2011.12.028
74. Trivizakis E, Souglakos J, Karantanas A, Marias K. Deep radiotranscriptomics of non-small cell lung carcinoma for assessing molecular and histology subtypes with a data-driven analysis. *Diagnostics (Basel)* 2021; **11**(12): 2383. https://doi.org/10.3390/diagnostics11122383
75. Park EK, Lee K-S, Seo BK, Cho KR, Woo OH, Son GS, et al. Machine learning approaches to Radiogenomics of breast cancer using low-dose perfusion computed tomography: predicting Prognostic biomarkers and molecular subtypes. *Sci Rep* 2019; **9**(1): 17847. https://doi.org/10.1038/s41598-019-54371-z
76. Aftimos P, Oliveira M, Irrthum A, Fumagalli D, Sotiriou C, Gal-Yam EN, et al. Genomic and transcriptomic analyses of breast cancer primaries and matched metastases in AURORA, the breast International group (BIG) molecular screening initiative. *Cancer Discov* 2021; **11**: 2796–2811. https://doi.org/10.1158/2159-8290.CD-20-1647
77. Luo Y, Cuneo KC, Lawrence TS, Matuszak MM, Dawson LA, Niraula D, et al. A human-in-the-loop based bayesian network approach to improve imbalanced radiation outcomes prediction for hepatocellular cancer patients with stereotactic body radiotherapy. *Front Oncol* 2022; **12**: 1061024. https://doi.org/10.3389/fonc.2022.1061024
78. Zhao L, Lee VHF, Ng MK, Yan H, Bijlsma MF. Molecular subtyping of cancer: Current status and moving toward clinical applications. *Brief Bioinform* 2019; **20**:

572–84. https://doi.org/10.1093/bib/bby026
79. Luo Y, McShan D, Ray D, Matuszak M, Jolly S, Lawrence T, et al. Development of a fully cross-validated bayesian network approach for local control prediction in lung cancer. *IEEE Trans Radiat Plasma Med Sci* 2019; **3**: 232–41. https://doi.org/10.1109/TRPMS.2018.2832609
80. Tseng HH, Luo Y, Cui S, Chien JT, Ten Haken RK, Naqa IE. Deep reinforcement learning for automated radiation adaptation in lung cancer. *Med Phys* 2017; **44**: 6690–6705. https://doi.org/10.1002/mp.12625
81. Niraula D, Sun W, Jin J, Dinov ID, Cuneo KC, Jamaluddin J, et al. A decision support software for AI-assisted decision making in response-adaptive radiotherapy — an evaluation study. *International Journal of Radiation Oncology*Biology*Physics* 2022; **114**: e101–2. https://doi.org/10.1016/j.ijrobp.2022.07.895
82. Niraula D, Jamaluddin J, Matuszak MM, Haken RKT, Naqa IE. Quantum deep reinforcement learning for clinical decision support in oncology: application to adaptive radiotherapy. *Sci Rep* 2021; **11**(1): 23545. https://doi.org/10.1038/s41598-021-02910-y
83. Tarazona S, Balzano-Nogueira L, Gómez-Cabrero D, Schmidt A, Imhof A, Hankemeier T, et al. Harmonization of quality metrics and power calculation in multi-omic studies. *Nat Commun* 2020; **11**(1): 3092. https://doi.org/10.1038/s41467-020-16937-8
84. Mangul S, Martin LS, Hill BL, Lam AK-M, Distler MG, Zelikovsky A, et al. Systematic benchmarking of omics computational tools. *Nat Commun* 2019; **10**(1): 1393. https://doi.org/10.1038/s41467-019-09406-4
85. Weber LM, Saelens W, Cannoodt R, Soneson C, Hapfelmeier A, Gardner PP, et al. Essential guidelines for computational method benchmarking. *Genome Biol* 2019; **20**(1): 125. https://doi.org/10.1186/s13059-019-1738-8
86. Morin O, Vallières M, Braunstein S, Ginart JB, Upadhaya T, Woodruff HC, et al. An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. *Nat Cancer* 2021; **2**: 709–22. https://doi.org/10.1038/s43018-021-00236-2
87. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016; **3**(1): 160018. https://doi.org/10.1038/sdata.2016.18
88. Blum A, Wang P, Zenklusen JC. Snapshot: TCGA-analyzed tumors. *Cell* 2018; **173**: S0092-8674(18)30391-X. https://doi.org/10.1016/j.cell.2018.03.059
89. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The international cancer genome consortium data portal. *Nat Biotechnol* 2019; **37**: 367–69. https://doi.org/10.1038/s41587-019-0055-9
90. Therapeutically applicable research to generate effective treatments, phs000218. v1.p1. https://ocg.cancer.gov/programs/target
91. Perez-Riverol Y, Bai M, da Veiga Leprevost F, Squizzato S, Park YM, Haug K, et al. Discovering and linking public omics data SETS using the omics discovery index. *Nat Biotechnol* 2017; **35**: 406–9. https://doi.org/10.1038/nbt.3790
92. Foster ED, Deardorff A. Open science framework (OSF). *Jmla* 2017; **105**: 203–6. https://doi.org/10.5195/jmla.2017.88
93. Wang L, Xiao Y, Ping Y, Li J, Zhao H, Li F, et al. Integrating multi-omics for Uncovering the architecture of cross-talking pathways in breast cancer. *PLOS ONE* 2014; **9**(8): e104282. https://doi.org/10.1371/journal.pone.0104282
94. Reyes M, Meier R, Pereira S, Silva CA, Dahlweid F-M, von Tengg-Kobligk H, et al. On the Interpretability of artificial intelligence in radiology: Challenges and opportunities. *Radiol Artif Intell* 2020; **2**(3): e190043. https://doi.org/10.1148/ryai.2020190043
95. Luo Y, Tseng H-H, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR Open* 2019; **1**: 20190021: 20190021: . https://doi.org/10.1259/bjro.20190021
96. Carvalho DV, Pereira EM, Cardoso JS. Machine learning interpretability: A survey on methods and metrics. *Electronics* 2019; **8**: 832. https://doi.org/10.3390/electronics8080832
97. Koh PW, Liang P. Understanding black-box predictions via influence functions. In: Precup D, Teh YW, eds. Proceedings of the 34th International Conference on Machine Learning. Vol 70. Proceedings of Machine Learning Research. PMLR; 2017:1885-1894. https://proceedings.mlr.press/v70/koh17a.html
98. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE* 2015; **10**(7): e0130140. https://doi.org/10.1371/journal.pone.0130140
99. Shrikumar A, Greenside P, Kundaje A. Learning Ig important features through propagating activation differences. In: Precup D, Teh YW, eds. Proceedings of the 34th International Conference on Machine Learning. Vol 70. Proceedings of Machine Learning Research. PMLR; 2017:3145-3153. https://proceedings.mlr.press/v70/shrikumar17a.html
100. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In: Precup D, Teh YW, eds. Proceedings of the 34th International Conference on Machine Learning. Vol 70. Proceedings of Machine Learning Research. PMLR; 2017:3319-3328. https://proceedings.mlr.press/v70/sundararajan17a.html
101. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med* 2018; **15**: e1002683. https://doi.org/10.1371/journal.pmed.1002683
102. Bibault J-E, Hancock S, Buyyounouski MK, Bagshaw H, Leppert JT, Liao JC, et al. Development and validation of an interpretable artificial intelligence model to predict 10-year prostate cancer mortality. *Cancers* 2021; **13**: 3064. https://doi.org/10.3390/cancers13123064
103. Hong JC, Eclov NCW, Dalal NH, Thomas SM, Stephens SJ, Malicki M, et al. System for high-intensity evaluation during radiation therapy (SHIELD-RT): A prospective randomized study of machine learning–directed clinical evaluations during radiation and chemoradiation. *JCO* 2020; **38**: 3652–61. https://doi.org/10.1200/JCO.20.01688
104. McIntosh C, Conroy L, Tjong MC, Craig T, Bayley A, Catton C, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat Med* 2021; **27**: 999–1005. https://doi.org/10.1038/s41591-021-01359-w