ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ANNALS OF ONCOLOGY driving innovation in oncology

REVIEW

Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review

A. Prelaj[1,2,3*], V. Miskovic[1,2], M. Zanitti[4], F. Trovo[2], C. Genova[5,6], G. Viscardi[7], S. E. Rebuzzi[6,8], L. Mazzeo[1,2], L. Provenzano[1], S. Kosta[4], M. Favali[2], A. Spagnoletti[1], L. Castelo-Branco[9,10], J. Dolezal[11], A. T. Pearson[11], G. Lo Russo[1], C. Proto[1], M. Ganzinelli[1], C. Giani[1], E. Ambrosini[2], S. Turajlic[12], L. Au[13,14,15], M. Koopman[16,3], S. Delaloge[17,3], J. N. Kather[18], F. de Braud[1], M. C. Garassino[11], G. Pentheroudakis[9], C. Spencer[12*†] & A. L. G. Pedrocchi[2†]

[1]Medical Oncology Department, Fondazione IRCCS Istituto Nazionale Tumori, Milan; [2]Nearlab, Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milano, Italy; [3]ESMO Real World Data and Digital Health Working Group, ESMO, Lugano, Switzerland; [4]Department of Electronic Systems, Aalborg University Copenhagen, Denmark; [5]UO Clinica di Oncologia Medica, IRCCS Ospedale Policlinico San Martino, Genoa; [6]Department of Internal Medicine and Medical Specialties (Di.M.I.), University of Genoa, Genoa; [7]Precision Medicine Department, Università degli Studi della Campania Luigi Vanvitelli, Naples; [8]Medical Oncology Unit, Ospedale San Paolo, Savona, Italy; [9]ESMO European Society for Medical Oncology, Lugano, Switzerland; [10]NOVA National School of Public Health, Lisboa, Portugal; [11]Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, USA; [12]Cancer Dynamics Laboratory, The Francis Crick Institute, London; [13]Renal and Skin Unit, The Royal Marsden NHS Foundation Trust, London, UK; [14]Department of Medical Oncology, Peter MacCallum Cancer Centre, Melbourne; [15]Sir Peter MacCallum Department of Medical Oncology, The University of Melbourne, Melbourne, Australia; [16]Department of Research and Development, Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands; [17]Department of Cancer Medicine, Gustave Roussy, Villejuif, France; [18]Else Kroener Fresenius Center for Digital Health, Medical Faculty Carl Gustav Carus, Technical University Dresden, Dresden, Germany

Available online 23 October 2023

Background: The widespread use of immune checkpoint inhibitors (ICIs) has revolutionised treatment of multiple cancer types. However, selecting patients who may benefit from ICI remains challenging. Artificial intelligence (AI) approaches allow exploitation of high-dimension oncological data in research and development of precision immuno-oncology.

Materials and methods: We conducted a systematic literature review of peer-reviewed original articles studying the ICI efficacy prediction in cancer patients across five data modalities: genomics (including genomics, transcriptomics, and epigenomics), radiomics, digital pathology (pathomics), and real-world and multimodality data.

Results: A total of 90 studies were included in this systematic review, with 80% published in 2021-2022. Among them, 37 studies included genomic, 20 radiomic, 8 pathomic, 20 real-world, and 5 multimodal data. Standard machine learning (ML) methods were used in 72% of studies, deep learning (DL) methods in 22%, and both in 6%. The most frequently studied cancer type was non-small-cell lung cancer (36%), followed by melanoma (16%), while 25% included pan-cancer studies. No prospective study design incorporated AI-based methodologies from the outset; rather, all implemented AI as a *post hoc* analysis. Novel biomarkers for ICI in radiomics and pathomics were identified using AI approaches, and molecular biomarkers have expanded past genomics into transcriptomics and epigenomics. Finally, complex algorithms and new types of AI-based markers, such as meta-biomarkers, are emerging by integrating multimodal/multi-omics data.

Conclusion: AI-based methods have expanded the horizon for biomarker discovery, demonstrating the power of integrating multimodal data from existing datasets to discover new meta-biomarkers. While most of the included studies showed promise for AI-based prediction of benefit from immunotherapy, none provided high-level evidence for immediate practice change. A priori planned prospective trial designs are needed to cover all lifecycle steps of these software biomarkers, from development and validation to integration into clinical practice.

Key words: immunotherapy, artificial intelligence, multiomics, real-world, multimodal

*Correspondence to: Dr Arsela Prelaj, Medical Oncology Department 1, Fondazione IRCCS Istituto Nazionale Tumori, Via Giacomo Venezian, 1, 20133 Milan, Italy. Tel: +39-0223903647

E-mail: arsela.prelaj@istitutotumori.mi.it (A. Prelaj).

*Dr Charlotte Spencer, Cancer Dynamics Laboratory, The Francis Crick Institute, 1 Midland Road, London NW11AT, UK.

E-mail: Charlotte.spencer@crick.ac.uk (C. Spencer).

†These authors contributed equally.

INTRODUCTION

The advent of immunotherapy (IO), more specifically immune checkpoint inhibitors (ICIs), has transformed management of many patients with cancer, including melanoma,[1] head and neck cancer (HNC),[2] bladder,[3,4] kidney,[5] and advanced non-small-cell lung cancer (NSCLC).[6-9] However, there are limitations with this approach; clinical

benefit is only achieved in a subgroup of patients, many challenges for clinical practice adoption and indications have been reported by experienced physicians, and treatment efficacy differs according to tumour type.[10] Accurate patient selection for specific therapies remains a crucial unmet clinical need. Few biomarkers are validated for clinical use and most applicable only to certain cancers such as programmed death-ligand 1 (PD-L1)[11] and microsatellite instability (MSI)[12] in lung and colorectal cancers, respectively. Instead, in other cancers patients currently qualify for ICI without biomarkers to predict response, meaning many patients suffer toxicity without deriving benefit. Given the complexity of the tumour microenvironment (TME) and immune system (innate and adaptive), it is unlikely that a single biomarker will be identified to robustly profile prognosis and prediction alone.[13] Instead, artificial intelligence (AI)-based methods promise an approach to define novel meta-biomarkers by integrating the multi-omics datasets now available in oncology [genomics, pathomics, radiomics, TME heterogeneity, and from more real-world data (RWD) generation], which are too large and complex for standard analytical tools.[14] We systematically reviewed published studies that leverage AI to predict efficacy of ICI therapies in different cancer types, focusing on genomics, transcriptomics and epigenomics, radiomics, pathomics, RWD, and multimodal data. We aim to provide practising oncologists and clinical investigators a description of principles on AI uses and a landscape view of AI-based predictive biomarkers for benefit from immunotherapy based on a systematic literature review, thus empowering an improved understanding of this rapidly evolving field. We also sought to highlight new potential meta-biomarkers for clinical consideration.

Important AI terminology used in this manuscript are presented in Table 1.

A subset of AI, machine learning (ML) is the array of techniques that learn from data and improve their performance iteratively to solve a specific task, using the available data about a phenomenon or a process. If the data consist of images, a standard ML models take as input a set of predefined features (e.g. tumour shape, tumour size) extracted from the data, not the data itself. In this case, feature extraction is not part of the learning process and is thus dependent on human expertise. However, the requirement of having a set of pre-defined hand-crafted features represents a significant limitation of standard ML techniques. To overcome such an issue, if a large enough amount of data are available, it is possible to use deep learning (DL), the branch of ML that instead exploits data in its raw format (i.e. DL can work directly with unstructured datasets, see Table 1) to discover and recognise patterns. DL makes predictions using multilayered neural network algorithms inspired by the neurological architecture of the brain. It has been shown also in other applicative fields that the use of DL techniques allows to achieve super-human performances in problems in which the hand-crafted features were not able to provide satisfactory results.[18] The most used DL method in analysing medical images is convolutional neural network (CNN) (see Table 1). Recently, vision transformers (ViT),[16] models that use self-attention mechanisms to process images, emerged as an alternative to CNN and are gaining increasing attention for the classification tasks in oncology.[19-21]

AI methodologies (standard ML and DL) can be broadly divided into (i) supervised, (ii) semi-supervised, and (iii) unsupervised learning. In supervised learning, the model learns from the labelled data, i.e. data with a known outcome. The two main types of supervised learning are classification and regression. Classification is used to predict categorical variables (e.g. whether the patient will respond to the treatment or not), while regression is used to estimate continuous variables, such as progression-free survival (PFS). Commonly used ML-supervised algorithms for both classification and regression in cancer research are random forest (RF)[22] and support vector machines (SVM).[23] Semi-supervised learning models deal with partially labelled datasets and thus are useful in situations where acquiring labelled data is time-consuming (for example, annotating a high number of medical images). The most common semi-supervised methods include multiple-instance learning (MIL)[24] and graph convolutional networks (GCNs).[25] Unsupervised learning is useful for discovering new patterns from the data, and it shifts to clustering and dimensionality reduction using algorithms such as principal component analysis (PCA) and k-means clustering. In oncology, survival analysis using the Cox proportional hazard model[26] is commonly used to identify the prognostic factors that have an impact on patients' recurrence or survival. Recently, several AI techniques, capable of accounting for interaction effects between features, have been adapted for this task.[27,28] The main advantage of these models is their success in handling censored data, frequently present in oncology. A frequently used AI algorithm for survival analysis is random survival forest (RSF).[29]

Despite the recent popularity of DL methods in cancer research, standard ML methods are used preferentially, especially in the context of structured data (see Table 1). Their main advantage is simplicity, which entails easier interpretability and transparency. Comparatively, DL model architectures are complex, comprising hundreds of layers and millions of parameters. The internal mechanisms of such architectures are not easily interpreted by a human, hence the training and prediction phases, and are often referred to as 'black box' approaches.[17] Explaining model decision-making strategy offers biological insight and has scientific discovery potential. Importantly it establishes trustworthiness of the models which is critical for efficient deployment to end-users—the clinicians and patients themselves.[30] The AI techniques used for explaining AI models and their predictions are known as explainable AI (XAI) methods.[17] Briefly, XAI methods can be divided into two broad categories: model-based and *post hoc*.[17,31] Model-based interpretability algorithms provide insights into the relationships between model parameters and the outcomes that they have learned (e.g. logistic regression), for example using mathematical approaches. *Post hoc*

Table 1. Glossary of terms

Artificial intelligence (AI)	**AI defines the ability of a computer to carry out tasks normally requiring human intelligence, such as image processing and decision making.**
Machine learning (ML)	ML is a subset of AI, based on algorithms that can learn from data, find patterns, and make interventions with minimal or without human intervention.
Artificial neural networks (ANN)	ANN are a subset of ML structurally and conceptually inspired by the human brain. They consist of input, hidden, and output layers with connected neurons (nodes). Although each node in a network can only carry out very simple computations, the network in total can solve complex pattern recognition tasks.
Deep learning (DL)	DL is a class of ML based on ANN having many hidden layers. DL can process data in their raw format (i.e. DL can work directly with unstructured datasets) to discover and recognise patterns.[15]
Convolutional neural network (CNN)	CNNs are a form of DL architectures, deep ANNs, mainly used to analyse images.
Vision transformer (ViT)	The ViT is a form of a visual model with architecture based on a transformer originally designed for natural language processing (NLP). ViT uses a self-attention mechanism to process images.[16]
Structured data	Data that are in a standardised format for organising and storing data, for example, a tabular data in an Excel file.
Unstructured data	Data that have no pre-defined data model and are not organised. Examples are images and videos as well as free text.
Area under the receiver operating characteristic curve (AUC-ROC)	Performance metric used in classification that represents the ability of a model to discriminate between two classes (e.g. responders versus non-responders). The higher the AUC-ROC (or AUROC) the better the model is at discriminating between classes. A score of 1 indicates perfect classification performance.
Sensitivity (or recall)	Evaluation metric used for classification: it answers the question of how sensitive the classifier is in detecting positive instances i.e. the true positive rate (true positive/true positive + false negative). It is considered as one of the most important in the medical field, since the goal is often to detect all positive instances.
Specificity	A metric that reflects the ability of a model to correctly identify negative instances i.e. the true negative rate (true negative/false positive + true negative).
Precision	Describes the fraction of true positives among all the samples predicted to be positive i.e. the positive predictive value = true positive/true positive + false positive.
F1 score	It is often referred to as the harmonic mean between precision and sensitivity, since it penalises extreme values in both cases. The higher the F1 score the model performs better.
Accuracy	An evaluation metric used for classification reflecting the ratio between the correctly classified samples and the total number of samples. In case of an imbalanced dataset it can be misleading, and in this case it is recommended to use other metrics, such as F1 score, sensitivity, and specificity. The higher the accuracy the better the model performs.
Concordance-index (C-index)	Performance metric used in survival analysis, which measures the ability of a model to separate censored data. Similar to AUC-ROC, the higher C-index the better the model is. A C-index of 1 indicates perfect predictive accuracy.
eXplainable artificial intelligence (XAI)	XAI aims at explaining, justifying, and understanding the decisions that are made by AI models.[17]
Multimodal data	Multimodal data refers to datasets that include different modalities, e.g. structured data and images.[13]
Supervised learning	In supervised learning the algorithm is trained on a set of data that includes both the input data (e.g. gene expressions, real-world data) and the desired output (e.g. whether a tumour is benign or malignant). The algorithm learns to identify patterns in the data that are associated with the set of observed output, becoming a prediction tool. Supervised learning includes classification and regression algorithms.
Unsupervised learning	In unsupervised learning, the model is trained to find patterns in an unlabelled dataset. Unsupervised learning includes clustering and dimensionality reduction approaches.
Semi-supervised learning	Semi-supervised approaches work on both labelled and unlabelled data to train a model. It is often used when there is not enough labelled data available to train a standard supervised model.
High-dimensional data	High-dimensional data in which the number of features is larger than the number of included instances (e.g. patients).
Meta-biomarker	Novel biomarkers derived from AI integration of multimodal data, i.e. new fused biomarkers different from individual unimodal biomarkers.

explainability targets more complex models that are not interpretable by their design and attempt to extract from the model implementation, e.g. through examples, the relationship provided by the model between input data and final outputs (e.g. saliency mapping[32]).

In this review, we focus on the application of AI methodologies to predict the incidence of positive outcomes in patients undergoing ICI therapies, following prior training of the model on accumulated big data. Figure 1 presents the main steps for the development of such a system, and illustrates an end-to-end pipeline operating on different data types for ICI modelling.

MATERIALS AND METHODS

This is a systematic review conducted following the PRISMA guidelines (Supplementary Information 1, available at https://doi.org/10.1016/j.annonc.2023.10.125).[33]

Search methods

A systematic search was done (EA) in October 2022 in the following bibliographic electronic databases: PubMed, Scopus, Web of Science, and Cochrane Library using the keywords listed in Supplementary Information 2, Table S1, available at https://doi.org/10.1016/j.annonc.2023.10.125**.** Keywords are subdivided into four categories: (i) data type, (ii) treatment, (iii) condition, and (iv) methodology used for the analysis. The keywords were combined using the Boolean operators (AND/OR) as follows: (1 OR 2 OR ... 13) AND (14 OR 15 OR ... 26) AND (27 OR 28 OR ... 31) AND (32 OR 33 OR … 37). We restricted the research to those papers published between 1991 and 1 October 2022 and in the English language.

Data selection and analysis

All identified studies were managed by using Mendeley Reference Management and Excel (Supplementary

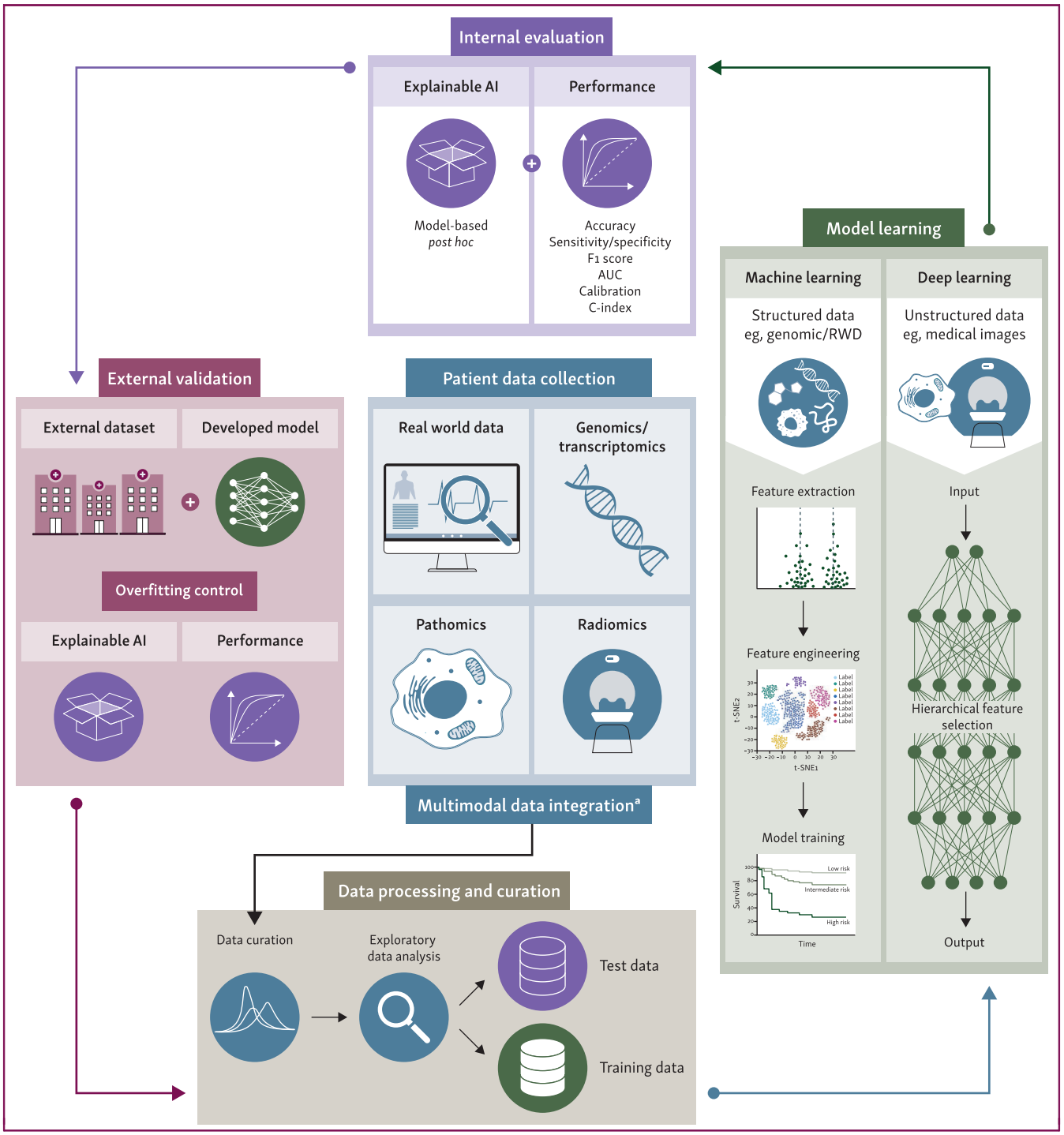

Figure 1. General steps in developing a model to predict IO treatment effectiveness, showing commonly used methodologies for different data types.
The workflow comprises three main steps. Step 1: Data processing and curation—omics data and/or clinical data and/or images are appropriately collected and stored integrated (when applicable) and pre-processed. The processed data are divided into a training and a testing dataset. Step 2: Model learning: Different techniques can be applied for the model to learn from the training dataset. Standard ML is a suitable choice if the data types are structured (for instance, RW and genomic data), while deep learning is used mainly for images (digital pathology and radiomics). The learning approach (supervised, semi-supervised, unsupervised) is steered by the end goals and the availability of labelled data. Steps 4 and 5: Internal and external validation: The trained model's performance is evaluated on the test dataset, which holds the "'ground truth'. Parallelly, how the model yields the prediction is explained. The predictive power and explainability of the model are validated on external datasets to assess its robustness and generalizability on unseen data (e.g. data from a different medical centre). Depending on the results of internal and external evaluation, new hypotheses can be formulated to refine data collection and train improved models.
[a]Multimodal data integration can be carried out in different stages of the pipeline.
AI, artificial intelligence; AUC, area under the curve; ML, machine learning; RWD, real-world data.

Information 3, available at https://doi.org/10.1016/j.annonc.2023.10.125). Duplicated studies were removed, and the full text of the articles was searched manually and using the Mendeley software. Two independent investigators (AP, VM) assessed the records after removing the duplicates at the level of title and abstract, and finally at

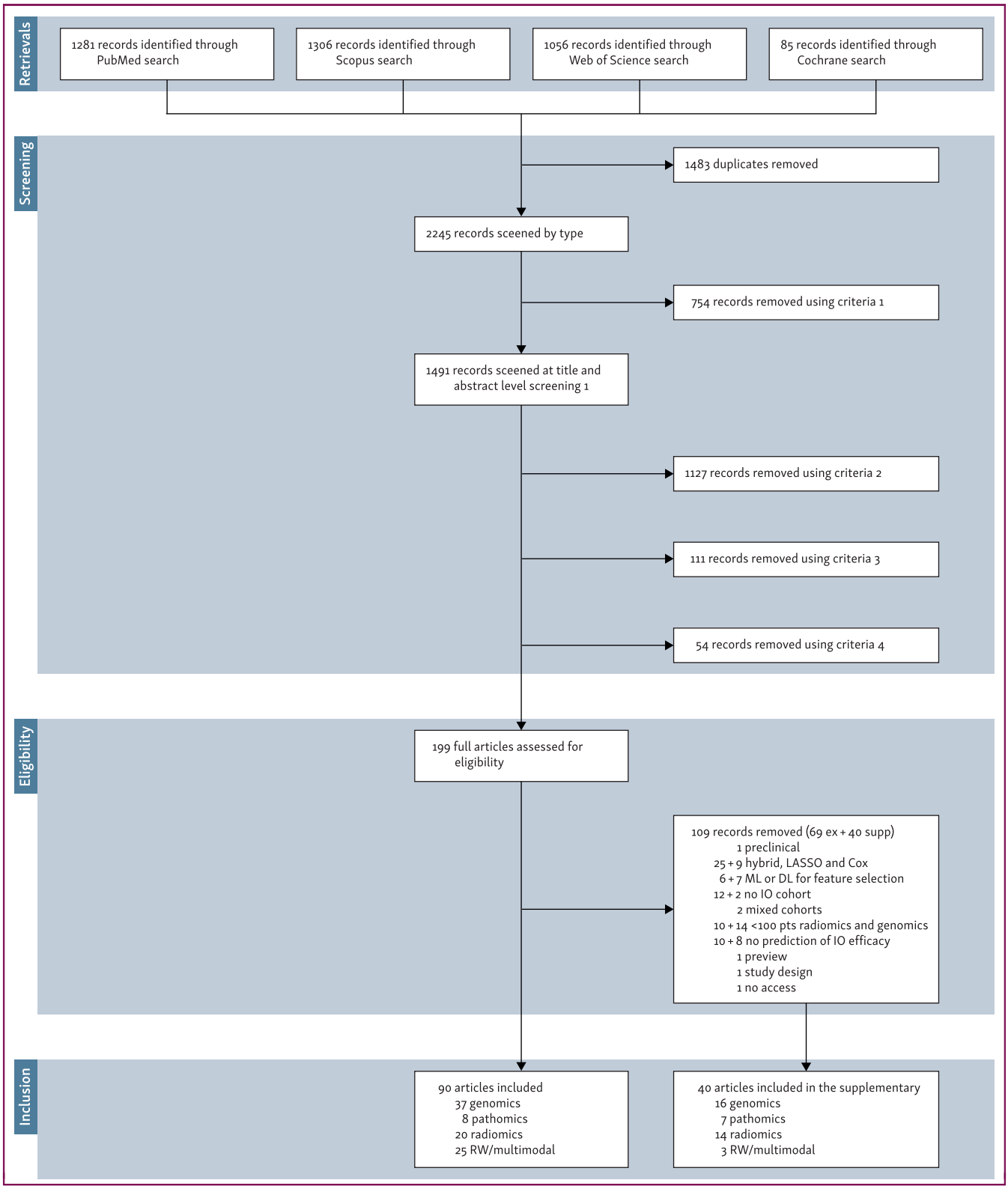

Figure 2. PRISMA flow diagram for study. In total, 3728 records were identified from a database search.
After removal of the duplicates and screening the records by types, we completed the screening of titles and abstracts for 1491 records and full-text screening for 199 records, 110 of which did not meet our inclusion criteria. 90 studies were included in the main text; 37 (41%) relate to genomics, 8 (9%) to pathomics, 20 (22%) to radiomics, and 25 (28%) to real-world and multimodal data.
DL, deep learning; IO, immunotherapy; ML, machine learning; RW, real-world.

the level of the full text, according to the inclusion and exclusion criteria. A detailed explanation of inclusion and exclusion criteria is provided in Supplementary Information 2, Table S2, available at https://doi.org/10.1016/j.annonc.2023.10.125. In short, the review included the following:

- Original peer-reviewed research published in journals
- Studies using ML and DL algorithms for the prediction of ICI efficacy in cancer patients
- Studies with an IO cohort as training and/or testing and/or validation set
- For genomics and radiomics data, studies with more than 100 patients

If any disagreement occurred, a third investigator (FT) with expertise in the field was contacted.

Data extraction

Data were extracted by two independent investigators (one with clinical and another one with technical background) for each data modality based on the division of the selected studies on (i) genomics and transcriptomics and epigenomic (GV, LP), (ii) radiomics (SER and MF), (iii) pathomics (CG, MG), and (iv) real-world and multimodal data (LM, MZ). Finally, two other independent investigators (AP, VM) revised all extracted data. We considered multimodal data to integrate three or more data modalities, while any studies with two modalities where one is RWD were not considered as multimodal. To provide a comprehensive summary of the identified studies, we investigated the following segments of the included studies: (i) type of the study (retrospective/prospective), (ii) data sources (public databases/clinical trials/institutions), (iii) cancer type, (iv) number of patients, (v) type of therapy, (vi) significant biomarkers, (vii) feature selection methodology (if applicable), (viii) developed model, and (xi) achieved results. Results for each section are presented in the form of tables. If the study included cohorts other than ICI cohorts, the tables report results related to just the ICI cohort.

Objectives

Firstly, we aimed to identify and appraise state-of-the-art AI methodologies used to predict response to ICI treatment in studies across four major data modalities: genomics/transcriptomics/epigenomics, radiological, histopathological images, and RWD/multimodal data. Secondly, we aimed to analyse the added value of AI methods in order to identify biomarkers or meta-biomarkers as predictors of ICI efficacy across different types of cancers.

RESULTS

Identified studies

Figure 2 summarises the study selection process. In total, 90 studies were eligible for final analysis: However, 40 other studies (due to their quality, even they did not meet the final criteria) were selected (Supplementary Information 4, Table S3, available at https://doi.org/10.1016/j.annonc.2023.10.125) for the following reasons:

- Reason 1: studies that according to authors are using hybrid methodologies between ML methods and classical statistics, e.g. studies that are using just least absolute shrinkage and selection operator (LASSO) or Cox analysis ($n = 9$) and not using classical ML;
- Reason 2: radiomics or genomics studies that include <100 patients ($n = 14$);
- Reason 3: studies with no ICI cohort ($n = 2$);
- Reason 4: studies that are not predicting ICI efficacy but the correlation of two different models created from two different data modalities, i.e. radiomic signature predicting tumour-infiltrating lymphocytes (TILS) ($n = 8$); and
- Reason 5: studies that used ML or DL for feature selection but not for efficacy prediction ($n = 7$).

Almost all (98%) identified AI analyses are retrospective on the data retrieved from both retrospective or prospective observational study cohorts and randomised clinical trials patients' cohorts. Generally, the data were not collected and designed a priori to be analysed using AI methodologies.

Biomarkers across different data types

A large amount of biomarkers related to IO's efficacy can be extracted with AI techniques across a wide variety of different cancer types, demonstrating the general applicability of the models. Biomarkers can be grouped based on the data from which they are extracted, as described in Figure 3.

Genetics. An overview of the 37 identified studies using genomics, transcriptomics, and epigenomics data is provided in Table 2. Among them, 24% ($n = 9$) are using genomics, 60% ($n = 22$) transcriptomics and 16% ($n = 6$) epigenomics data, and 65% ($n = 24$) of selected articles were published in 2022. In Supplementary Information 4, Table S3, available at https://doi.org/10.1016/j.annonc.2023.10.125, we included 16 studies that used genomics and transcriptomics data but did not comply with the inclusion criteria. The most studied single type of cancer was NSCLC (22%, $n = 8$), followed by melanoma (11%, $n = 4$) and bladder cancer (8%, $n = 3$). However, most studies (35%, $n = 13$) used multiple cancer cohorts. We identified two scenarios to build predictive AI models in the selected studies: (i) standard ML or DL models to directly predict patient outcome (48%), (ii) standard ML or DL models to create a genomic signature validated by statistical analysis to stratify patients into low- and high-risk groups (52%). Seventy-three percent ($n = 27$) of the studies used datasets from publicly available data platforms such as The Cancer Genome Atlas Program (TCGA),[71] Gene Expression Omnibus (GEO),[72] and cbioportal.[73,74]

Genomics. To generate large volumes of data, crucial for the successful implementation of AI methods, 35% ($n = 13$) of all genetic studies used multiple cancer types for the model

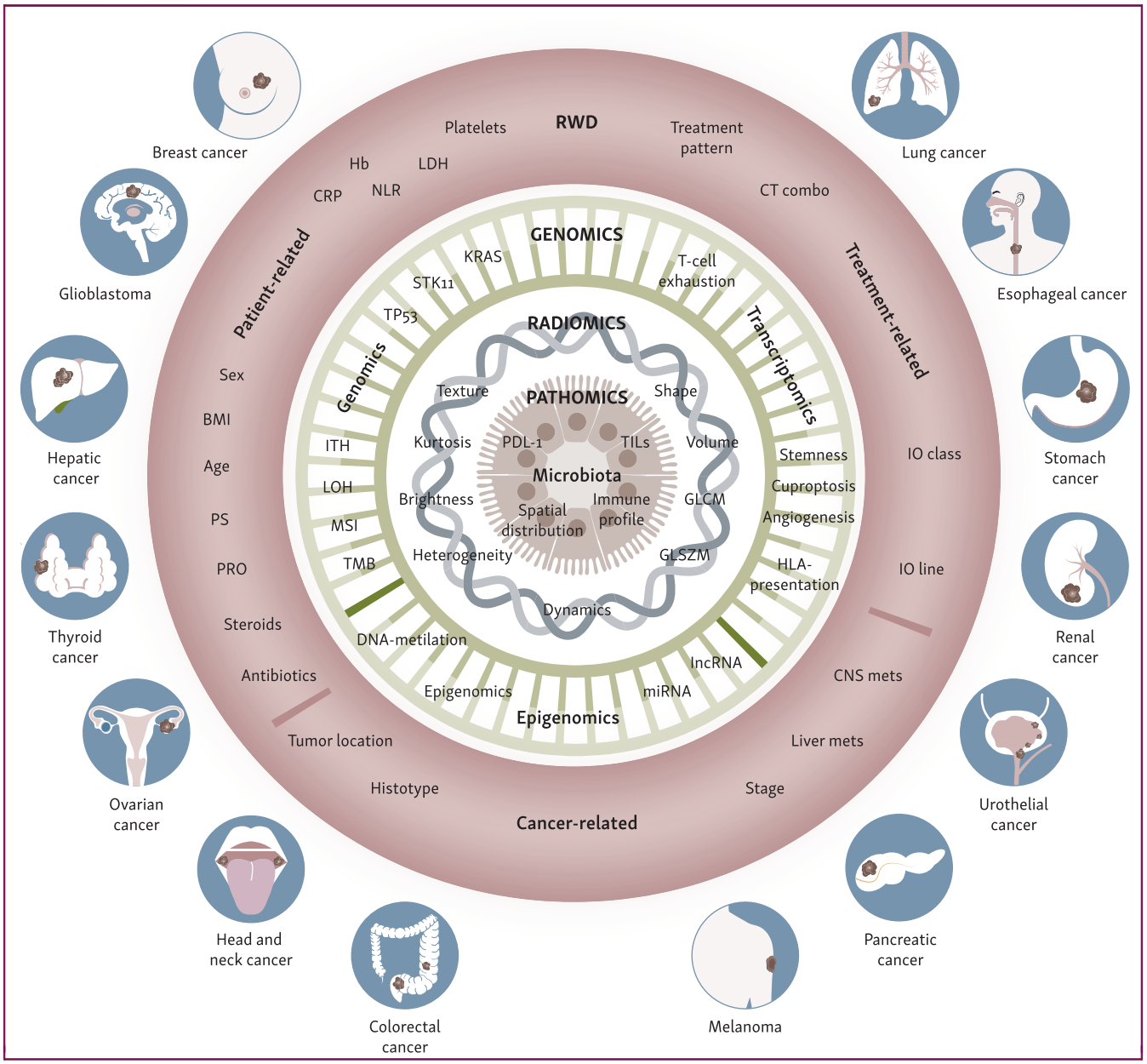

Figure 3. Relevant biomarkers emerged from AI-based predictive algorithms.
Biomarkers identified in the selected studies are grouped in layers organised on the criterion of the multiple data sources analysed using ML methods. In the most external layer of the sphere, features from RWD are clustered in patient-, cancer-, and treatment-related. The 'genomics' layer includes biomarkers from genomics, epigenomics, and transcriptomics; the 'radiomics' layer encompasses first-order and shape-based statistics, matrices, and delta-radiomics. The core of the sphere, lastly, contains biomarkers from pathomics and microbiota. BMI, body mass index; CNS, central nervous system; CRP, C-reactive protein; CT, chemotherapy; GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix; HLA, human leukocyte antigen; IO, immunotherapy; ITH, intratumour heterogeneity; LDH, lactate dehydrogenase; lncRNA, long non-coding RNA; LOH, loss of heterozygosity; miRNA, microRNA; MSI, microsatellite instability; NLR, neutrophils-to-lymphocyte ratio; PD-L1, programmed death-ligand 1; PRO, patient reported outcome; PS, performance status; RWD, real-world data; TILs, tumour-infiltrating lymphocytes; TMB, tumour mutational burden.

development. For example, Chowell et al.[34] developed an RF classifier using 16 genomic features across 16 cancer types to predict response to IO. The developed model discriminated responders (R) from non-responders with an area under the curve (AUC) of 0.79, significantly outperforming tumour mutational burden (TMB), a Food and Drug Administration-approved biomarker for this purpose,[75] in predicting overall survival (OS) ($P < 0.0001$) and PFS ($P < 0.0001$). There were different approaches to handle high-dimensional data (see Table 1). Most studies (92%, $n = 34$) used feature selection methods, divided mainly into clinician-driven hypothesis (clinical expertise, available literature, 18%, $n = 6$), classical statistics (feature correlation, univariate analysis 15%, $n = 5$), and automatic and data-driven feature selection ML techniques [Gene Set Enrichment Analysis (GSEA), LASSO, univariate Cox analysis, RF, SVM, minimum redundancy maximum relevance, and other ML models, 68%, $n = 23$]. Additionally, the network-based computational approach[76] showed potential in solving this problem.[77] For example, Kong et al.[35] used it to create an ICI treatment biomarker in combination with logistic regression (LR) achieving robust prediction in

https://doi.org/10.1016/j.annonc.2023.10.125

Table 2. Studies using genetics features and artificial intelligence to predict immunotherapy response in cancers

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Genomics									
Chowell et al., 2022[34]	Retrospective	Institutional (MSKCC)	NSCLC (37%), melanoma (13%), renal, bladder, head and neck, sarcoma, endometrial, gastric	1479 (total) 1184 (train) 295 (test)	Anti-PD-1/PD-L1 ± anti-CTLA-4	TMB, chemo, prior ICI, albumin, NLR, age, platelets, FCN, BMI, HED, Hb, cancer type, LOH, sex, drug class, tumour stage, MSI status	Literature, clinical	ML; classification; RF (with different number of features 11 and 16) Survival analysis; Statistics; K-M and CPH	Test set, pan-cancer (RF16/RF11) Responders versus non-responders AUC 0.79/0.71 OS C-index 0.68/0.62 HR = 0.29, $P < 0.0001$ PFS C-index 0.67/0.62 HR = 0.34, $P < 0.0001$
Kong et al., 2022[35]	Retrospective	TCGA, IMvigor210, other datasets	Melanoma, gastric and bladder cancer (T); melanoma (V)	729 (total) 605 (IO test) 37 (IO val1) 25 (IO val2) 49 (IO val3)	Anti-PD-1 /PD-L1 and /or anti-CTLA-4	Network-based biomarkers and TMB	SelectKBest, network-based approach	ML; classification; l2 regularised LR model; SVM, RF, DNN Survival analysis (statistics); K-M, stratification using ML output	LR best model: responders versus non-responders: ex.val1 AUC 0.79 ex.val2 AUC 0.72 ex.val3 AUC 0.69
Peng et al., 2022[36]	Retrospective	Institutional (DFCI, MSKCC), other datasets	NSCLC	248 (total) 83 (train), 165 (test)	Anti-PD-1 /PD-L1 ± anti-CTLA-4	15-gene somatic mutations (SMS); TP53 mutation, PD-L1, TMB	LASSO	ML; classification; SVM, for development for SMS signature Survival analysis, (statistics): K-M, stratification between low and high SMS	Test set BOR: AUC 0.841; Sen. 91.27%; Spec. 61.67% PFS: HR 3.89, $P < 0.001$ OS: HR 2.82, $P < 0.001$
Fang et al., 2021[37]	Retrospective	FHFM-NSCLC database	NSCLC	1937 (total)	anti-PD-1/PD-L1 monotherapy or combined therapy	21-clinical genomic features; high and low blood monocyte count, KRAS mut	Literature, clinical	DL: deep patient graph convolutional network 'DeePaN', for DeePaN signature creation Survival analysis, (statistics): K-M, stratification between low and high DeePaN	OS: $P = 2.2 \times 10^{-11}$
Ahn et al., 2021[38]	Retrospective	Institutional (YCH)	NSCLC	192 (total) 142 (train), 50 (test)	Anti-PD-1 /PD-L1	NLR, PD-L1, TMB and number of metastases	Literature, clinical and correlation	ML; classification; XGBoost LightGBM, MNN, SVM, LDA, QDA, ridge regression and LASSO Survival analysis (statistics) K-M	Responders versus non-responders: Test set: LightGBM (best model) AUC 0.788

Continued

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Wang et al., 2020[39]	Retrospective	IMvigor210; TCGA-BLCA	Bladder cancer (development)	272 (for signature development and validation in survival analysis); 163 (train) 109 (test)) 406 (IO external validation)	ICI	TLS	SVM-RFE	ML; for TLS signature development: classification; ADA, k-NN, LASSO, MARS: NN, RF, SVMLinear, SVM Radical, XGBDART Survival analysis Statistics: K-M; satisfaction based on TLS score to low versus high TLS group	OS: Validation: HR 0.55, $P < 0.0001$ Ex validation: HR 0.70, $P = 0.031$
Peng et al., 2020[40]	Retrospective	cbioPortal; MSKCC, Van Allen database	Lung adenocarcinoma	405 (total) 179 (cbioPortal) 143 (train) 36 (test) 47 (ex. validation)	Anti-PD-1/PD-L1 ± anti-CTLA-4 (cbioPortal); anti-PD-1 (Van Allen)	Somatic mutations; STK11, EGFR and KMT2D mutations. Two mutational signatures (C1 and C2)	DNN during training	DL; classification; DNN—input layer, two hidden layers, a dropout layer, and an output layer Survival analysis (statistics); K-M; stratification based on model output DCB versus NDB	DCB versus NDB Test1: AUC 0.918, Sen. 89.47%, Spec. 94.12%. Ex. val: AUC 0.910, Sen. 93.33%, Spec. 90.63%. Test: OS HR = 0.335, $P = 0.039$; PFS HR = 0.215, $P < 0.0001$ Ex. val: OS HR = 0.128, $P = 0.00$; PFS HR = 0.240, $P < 0.0001$
Xie et al., 2022[41]	Retrospective	TCGA(T) others dataset(V1,V2)	Solid tumours (T) Melanoma(IO validation)	8818 (total) 8646 (training) 108 (IO validation1), 64 (IO validation2)	Ipilimumab	Four genomic classes model	Literature (i.e. TMB, MSI, SCNA)	DL for model generation (clustering): DBN to extract features of multifactorial inputs; DAE for clustering/stratification analysis)	IO val1: OS $P = 0.009$; ORR $P = 0.003$ IO combined: OS $P < 0.0001$ ORR $P = 0.0002$
Wang et al., 2022[42]	Retrospective	Institutional (TMUGH), other datasets	NSCLC	162 (total); 69 (training), 72 (test1), 21 (test2)	ICIs	multi-feature model based on genomic markers; TMB, ITH, and HLA LOH	Cross-validated recursive feature Elimination	ML; classification; k-NN, LR, SVM_rbf, SVM_ linear, SVM_poly, RF, GBC, DTC, ETC, and GPC Survival analysis (statistics); K-M; stratification based on model output DCB versus NDB	DCB versus NDB. SVM_poly. Test 1 AUC 0.77 Test 2 AUC 0.78 Test 1 OS HR = 2.16, $P = 0.0095$; PFS HR =1.97, $P = 0.086$ Test 2 PFS HR 11.25, $P = 0.0046$
Transcriptomics									
Wiesweg et al., 2020[43]	Retrospective	Institutional (WGCCES, HKEBHB)	NSCLC	122 (total) 55 (train); 36 (test); 31 (ex. validation)	Anti-PD-1 /PD-L1	RNA expression + cell type	LASSO	ML; regression and classification; SVM, RF, LR, GB	TTF (continues and categorised in two/three categories): Test HR 0.46, $P = 0.035$ Ex. val. and test combined: HR 0.4, $P = 0.033$

Continued

https://doi.org/10.1016/j.annonc.2023.10.125

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Charoentong et al., 2017[44]	Retrospective	TCGA; other datasets	20 solid tumours	8243 (total) 42 (IO validation1) 28 (IO validation2)	Anti-PD-1/PD-L1 and anti-CTLA-4	Immunophenoscore (panel of immune gene)	RF	ML for creating Immunophenoscore RF Survival analysis (statistics); K-M; CPH	IO val 1: AUC 1 (0.99-1) IO val 2: AUC 1 (0.99-1)
Ren et al., 2022[45]	Retrospective	TCGA, GEO	Lung adenocarcinoma	1906 (total) 298 (IO ex validation 1) 70 (IO ex. validation 2)	Not specified	TIME -multi-omics score	Clustering	ML: Unsupervised clustering PAM for signature creation Survival analysis (statistics) K-M, stratification on low and high time-score subgroups	OS Ex validation 1: $P < 0.001$ Ex validation 2: $P < 0.001$
Xu et al., 2021[46]	Retrospective	GEO (GSE91061 dataset-IO cohort)	Melanoma	1447 (total) IO cohort number not specified	anti-PD1 and anti-CTLA-4 therapy	ICS score	VIMP, RF, CPH	ML: for ICS score development ML: Classification XGBoost Survival analysis (statistics) K-M, stratifications on low and high ICS group	Train: Responders versus non-responders AUC = 0.926 OS: HR 2.28, $P = 0.0389$
Chen et al., 2022[47]	Retrospective	TCGA(T), CGGA(V1), IMvigor210 (IO cohort) XHACSU- (V2)	Diffuse glioma	NA(train) NA (test) NA (IO validation) 105 (validation 2)	anti-PD-L1 antibody atezolizumab	29-gene expression signature; IDH1, CIC, TP53, EGFR, ATRX	Clustering, elastic net regression, Cox	ML for signature creation: elastic regression analysis and PCA Model prediction: nomogram, and univariate and multivariate regression Survival analysis, CPH and K-M log-rank test	OS: stratification on low- versus high-risk score based on signature $P = 0.0022$
Zhang et al., 2022[48]	Retrospective	NCCJ	NSCLC	213 (total; fivefold cross-validation)	Nivolumab	45-miRNA expression and 3 clinical features-based model	ANOVA (responders versus non-responders)	ML; classification; RF (three classes)	Exceptional responders versus others AUC 0.764-0.830; Sen 0.63-0.71; Recall 0.4-0.81 F1 0.49-0.76
Ma et al., 2022[49]	Retrospective	TCGA, GEO, clinical trial (IMvigor210)	GC (Train) GC, Bladder carcinoma (test)	414 (total) 331 (train) 83 (test) 348 (IO validation)	Atezolizumab	36 AGR, significant mutation of FGFR4 and HER2	SVM, LASSO, RFB, XGBoost	ML for AGR score generation: MLR Survival analysis (statistics): K-M, stratifications on low and high AGR	Responders versus non-responders $P = 0.051$ OS: $P = 0.014$
Zhang et al.; 2022[50]	Retrospective	GEO (train) TCGA(V1), TISCH 10 others dataset (V2)	Melanoma (training) Mixed solid tumours (validation)	921 (total) 618 (train) 154 (validation) 149 (independent test)	Anti-PD-(L)1, anti-CTLA-4	Cancer stemness, Stem.Sig (RNA-expression-based signature)	Correlation	ML; classification SVM, NB, RF, k-NN, AdaBoost, LogitBoost, Cancerclass Survival analysis; K-M stratification on high- and low-risk groups	NB best model (results on independent test set) Responders versus non-responders AUC = 0.71 OS: $P = 2e-04$

Continued

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Liu et al., 2022[51]	Retrospective	TCGA, GEO, Institutional (FAHZ)	CRC(T) CRC, Bladder carcinoma (V)	1048 (total) 616 (train) 72 (validation) 56 (IO validation set1) 298 (IO validation set1)	Atezolizumab	Nine stemness-related gene expression signature: GFPT1, PTMAP9, MOGAT3, DPM3, S100A12, PGM5, FUT6, SEMA3C, ADAM33	LASSO, SVM, RF, XGB	ML: LR: for stemness cluster predictor Statistical analysis for comparing responders in two clusters	Responders to ICIs in two clusters IO val 1 $P = 0.01$ IO val 1 $P < 0.0001$
Chen et al., 2021[52]	Retrospective	Cbioportal (train, validation)	NSCLC	226 (total) 179 (train) 47 (ex. validation)	Anti-PD-(L)1, anti-CTLA-4	Six-gene genomic model ERBB4, ATRX, FAT1, KDM5C, ASXL1, and AR	LR, RF	ML, classification, RF, SVM, GLM Survival analysis for OS—CPH Classification for DCB—LR	DCB versus NDB RF (best model) AUC O.90 OS: $P = 0.00019$
Li et al., 2022[53]	Retrospective	TCGA(T) GEO(V1) GEO, clinical trial (IMvigor210- V2)	NSCLC (training, validation 1) Bladder, NSCLC (validation 2)	1771 (total) 492 (train) 925 + NA (validation 1) 325 (IO validation 2) 29 (patient with single-cell sequencing data)	Anti-PD-1	15 independent radiotherapy-related CRGs expression; MRPS16, EGFR, KSR1, LSM5, PES1, SAE1, SLBP, USP1; PTGES3, PRM1, SEMA3A, RAD51, ABCE1, COL4A6, TOMM40	Literature (CRGs list), GSEA (gene selection)	DL: ANN risk model, stratification on high- and low-risk groups TIDE analysis Survival analysis (statistics); K-M	OS difference between high- and low-risk groups (IO validation) $P < 0.0001$
Prelaj et al., 2022[54]	Retrospective	Institutional (APOLLO trial)	NSCLC	164 (total; fivefold cross-validation)	Nivolumab, pembrolizumab, atezolizumab	ECOG, NLR value, IO line, and MSC test level, PD-L1	Correlation	ML: classification: LR, FFNN, k-NN, SVM, RF	Responders versus non-responders Best-performing LR AIC = 132.5; Acc. = 0.756; F1 0.722; AUC 0.83
Zheng et al., 2022[55]	Retrospective	Bi's dataset, Au's dataset, GEO	ccRCC	182 (total) 10 (development) 172 (IO validation)	PD-1/PD-L1	ccRCC signature, 47 genes; Proliferative CD4+ T cells and Regulatory T cells and a subtype of antigen-presenting monocytes	GSVA	ML for validation developed ccRCC signature: classification: SVM, NB, KNN	Responders versus non-responders AUC between 0.85 and 1
Huang et al., 2022[56]	Retrospective	GEO, TCGA, clinical trial (IMvigor210)	HNSCC; NSCLC; melanoma; gastrointestinal cancer; urothelial carcinoma	566 (total; IO set) 501 (training) 65 (test)	anti-PD-1/PDL-1	Chemokines expression (CXCL9, CXCL10; CXCL11; CCL5) immune cells (macrophages M1, CD8 CD4 T cells)	GSVA scoring, correlation, statistical analysis, PCA and Umap for dimension reduction	DL: Classification four-layer neural network model	Responders versus non-responders: Training set AUC 95% Test set AUC 75%
Vathiotis et al., 2021[57]	Retrospective	Institutional (YCC)	Melanoma	59 (total) 80% (training) 20% (test set)	Nivolumab, pembrolizumab, ipilimumab	mRNA and spatially defined protein	Elastic net regularisation, univariate LR, univariate R-squared	ML: classification; elastic net regularised regression Survival analysis (statistics): Stratification in low- and high-risk groups using score	BOR; AUC 0.97; Sen. 0.96; Spec. 0.93; PPV 0.91; NPV 0.96 Low- versus high-risk groups: Median PFS HR 0.2, $P < 0.0001$ Median OS HR 0.16, $P < 0.0001$

Continued

https://doi.org/10.1016/j.annonc.2023.10.125

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Xu et al., 2022[58]	Retrospective	TCGA, clinical trial (IMvigor210 - IO cohort), GEO (IO cohort)	Bladder	1216 (total) 298 (IO test 1) 39 (IO test 2)	Not specified	30 robust OS-related genes XIAP, RPAIN, ARHGEF6, PRKAA1 LY6G5C, UBE2NL SLC6A13, RCOR1, NKAIN4, DCAF4, GPSM2, KLK10, SFTPD, DSC2, DSG3 DSG1, DSC3, SCEL, IMPA2, S100A7 SPRR2G, CDCA7L SRP68, SPRR1B, KRT4, LGALS7B ANXA2, IGFL, LGALS7, KRT6A	LASSO, stepwise Cox, CoxBoost, and RSF	ML for AIGS score generation: RSF, Enet, LASSO, Ridge, stepwise Cox, CoxBoost, plsRcox, SuperPC, GBM, and survival-SVM Statistical analysis, *t*-test responders versus non-responders based on ML-generated signature	Responders versus non-responders Test 1 $P < 0.001$ Test 2 $P < 0.001$
Kim et al., 2022[59]	Retrospective	Six cohorts from other datasets; 1 institutional (SMC)	NSCLC, melanoma	529 (total) Lung: 122 (institutional cohort) 75 (ex. cohort 1) 34 (ex cohort 2) Melanoma: 110 (ex cohort 3) 64 (ex cohort 4) 56 (ex cohort 5) 68 (ex cohort 6)	Anti-PD(L)-1 ICIs (NSCLC cohorts), nivolumab/ipilimumab (melanoma cohorts)	HLA-A, HLA-B, immune system, signalling by EGFR in cancer, signalling by EGFR, cytokine signalling in immune system, adaptive immune system.	Literature (immune epitope database), RF	DL: CNN: for prediction of neoantigen load ML: classification: RF (predictions of resistance to checkpoint blockade) Survival analysis: K-M, neoantigen load correlation with clinical benefit	OS: Lung cancer Inst. Val $P = 0.034$ Ex val1 $P = 0.001$ Ex val2 P 0.014 Melanoma: Ex val3 $P = 0.047$ Ex val4 P 0.002 AUC: Lung: Inst. Val 0.814 Ex val1 0.843 Ex val2 0.846 Melanoma: Ex val3 0.785 Ex val4 0.758 Ex val5 0.954 Ex val6 0.829
Liu et al., 2022[60]	Retrospective	TCGA; GEO	Melanoma	1290 (total) 454 (train) 561 (test) 275 (IO validation)	Anti-PD-1, anti-CTLA-4, anti-MAGE-A3	MLPS based on expression of 40 RSAGs	LASSO, stepwise Cox, CoxBoost, RSF	ML for signature creation: MLPS: stepwise Cox, CoxBoost, ridge regression, RSF, GBM, SVM, LASSO, SuperPC, plsRcox, Enet Survival analysis (statistics): K-M, log-rank	Ratio of responders in the high MLP group versus low MLPS group: 59.03 versus 20.7%; $P < 2.2e-16$

Continued

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Hu et al., 2022[61]	Retrospective	TCGA; GEO (validation), GEO (IO validation- GSE93157)	Gastric cancer (TGCA; GEO); NSCLC, HNSCC, melanoma (GSE93157)	1564 (total) 349 (TCGA); 964 (validation); 65 (IO validation1) 186 (IOvalidation2)	Anti-PD-1, anti-CTLA-4	nine eRNA pairs; three subtypes with distinct eRNA expression patterns	Clinical, univariant cox analysis, GO enrichment analysis	ML for eRNA signature creation: PCA ML for model generation: RF Statistics: TIDE algorithm, K-M, log-rank, Fisher's exact test, stratification on high and low eRNA score	IO validaion2 PFS $P = 0.0072$ Improved response high eRNA versus low eRNA $P < 0.05$ IO validaion2 PFS $P < 0.0001$ Improved response high eRNA versus low eRNA $P = 0.0082$
Wang et al., 2021[62]	Retrospective	TCGA; CGGA; institutional PUMCH (validation)	Glioblastoma	906 (total) 518 (train); 350 (validation) 38 (validation)	Anti-PD-1, anti-CTLA-4	Seven critical genes selected by all algorithms: LTF, TAGLN, C5AR1, RAB33A, CFI, CH24H, RNASE2	LASSO, SVM, RFB, XGB	ML for creation of predictive model 'stemness subtype predictor'; multivariant LR Statistics: K-M, log-rank test	Responders in the stemness subtype I group versus the stemness subtype II group $P < 0.001$
Zhang et al., 2022[63]	Retrospective	TCGA (test and training); GEO others dataset (validation)	Mixed solid tumours (TCGA); melanoma, RCC, GBM, UC (validation)	9564 (total) (training); 273 (IO validation)	Anti-PD-1/PD-L1 or anti-CTLA-4	ML-TEX gene signature score: TNF, IL2, IFN-γ, CTL, TGF-β, IL-10, glycolysis, chemokines, CYT, CD8+ T cell, CD4+ T cell, Th1, Th2, Th17, TCF1, T-bet, TOX, TCR, BCR TEX-related cancer driver genes in the IO cohort: LTB, TLL1, HLA-B, SFMBT2, LOX, IKZF3, TCF4, PDGFRB, ZNF521, PRF1, ZEB1, IRF4, CDH11, MET, MYH11, AXN2, B2M, JAK1, NFKBIE, P2RY8, DCC, SP140, LATS2, CD79B, CR1, HLA_A, EML4, IL7R, BTK, CCR7, PSDH17, RBM10, ERG, MSI2, CBL, DDB2, PRKD2, WAS, STAT6, IRF1	Correlation, deep neural network	ML for TEX score generation (TEXprog, TEXint1 and TEXint2, TEXint3 and TEXterm) XGBoost, multi-logistic, RF, SVM, FFNN Survival analysis (statistics): K-M, log-rank, stratification on responders and non-responders using TEX	Responders versus non-responders in TEX^{int2} versus other four TEX groups: $P < 0.001$ OS: $P < 0.05$ in all different IO cohorts

Continued

https://doi.org/10.1016/j.annonc.2023.10.125

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Liu et al., 2022[64]	Retrospective	GEO	PDAC, BCC, Melanoma	62 (total), 19 (PDAC) 11 (BCC) 32 (melanoma) 32 80% (training) 20% (test)	Anti-PD-1	MHC expression, CD8+ cells	Clinical	ML Classification: MLP neural network	Responder/no responder: Test set BCC: Acc. 96.7% Melanoma: Acc. 60.7%
Epigenomics									
Zhang et al., 2022[65]	Retrospective	TCGA, CGGA, GEO, 20 datasets (e.g. IMvigor210), institutional (XHACSU)	LGG, Mixed solid tumours	932 (total) In total 10 IO validation cohorts, number not specified	Not specified ICIs	TIIClncRNA-based signature	101 combination of 10 ML algorithms (e.g. LASSO, RSF, CoxBoost, GBM)	ML; for TIIClncRNA score generation; RSF and CoxBoost Survival analysis (statistics) K-M, stratification on low TIIClncRNA versus high TIIClncRNA	Responders versus non-responders or OS $P < 0.05$ for all IO validation cohorts
Filipski et al., 2021[66]	Retrospective	TCGA (model building); institutional validation, (GUF, CUB, UHW)	Melanoma	531 (total) 470 (TGCA); 61 (IO cohort)	ICIs	CpG sites, specific for LMC Significant: Tregs, CD56+, fibroblasts	MeDeCom algorithm	ML: Clustering and Classification: LASSO Survival analysis: K-M (statistics)	OS (cluster 1 versus cluster 2) $P = 0.042$ Good versus poor survival AUC 0.9664 Acc. 89.5%
Xu et al., 2021[67]	Retrospective	TCGA (model building); other datasets (validation)	32 tumour types (TCGA); NSCLC (validation)	7209 (total) 7131 (TCGA); 78 (IO validation)	Anti-PD-1/PD-L1	DNA methylation profile	Based on prior knowledge (creation of two clusters based on immune infiltration analysis)	ML for model generation: *k*-NN, RF, LR, SVM Survival analysis, log-rank test	SVM (best model) Responders versus non-responders (validation) F1 0.4255; MCC 0.18993; AUC 0.6742 PFS $P = 0.06$
Zhou et al., 2021[68]	Retrospective	GEO, TCGA	Bladder cancer, ccRCC (IO validation)	403 (total) 202 (training) 201 (validation) 151 (ex validation) 16 (IO ex. validation)	Anti-PD-1 nivolumab	TILBlncSig	GSEA, Cox regression models and a forward and backward variable selection	ML: for TIL-B-derived lncRNA signature creation: multivariate Cox regression models and a forward and backward variable selection Survival analysis (statistics), K-M, log-rank test, stratification on high- and low-risk groups	OS: $P = 0.098$ AUC 0.719

Continued

Table 2. Continued

Authors and years	Type of analysis	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Liu et al., 2022[69]	Retrospective	TCGA; GEO; more public dataset; Institutional (FAHZ - IO cohort)	CRC	816 (total) 584 (training) 232 (IO validation cohort)	Pembrolizumab	IRLS signature (43 lncRNAs)	GSEA, PAC	ML: Survival analysis for IRLS signature development: 101 prediction models, combination of: RSF, Enet, LASSO, Ridge, stepwise COX, CoxBoost, plsRcox, SuperPC, GBM, SVM Survival analysis (statistics): K-M; log-rank	LASSO + stepCox (best model) OS: $P < 0.0001$ RFS: $P < 0.0001$ Time-dependent AUC: 1-year 0.840 3-year 0.776 5-year 0.818 C-index 0.765
Pan et al., 2022[70]	Retrospective	TCGA (development); and GEO-GSE (validation)	Lung adenocarcinoma and NSCLC (IO cohort)	1225 (total) 437 (train) 60 (IO validation1) 18 (IO validation2)	anti-PD-1 /PD-L1	Epigenome signature; iDMCs	mRMR	ML: for creating signature iPMS; RF Survival analysis, Cox regression, K-M, log-rank	NDB versus DCB IO val. 1 AUC = 0.752 IO val. 2 AUC = 0.653

Acc, accuracy; AdaBoost, adaptive boosting; AGR, angiogenesis-related gene expression signature; AIC, Akaike information criterion; AIGS, artificial intelligence-derived gene signature; ANN, artificial neural network; ANOVA, analysis of variance; ATRX, alpha thalassemia/mental retardation X-linked; AUC-ROC, area under the receiver operating characteristic curve; BLCA, bladder cancer; BMI, body mass index; BOR, best overall survival; C1, consensus clusters 1; ccRCC, clear cell renal cancer carcinoma; CD56+, cluster of differentiation56; CGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas, TMB, tumour mutational burden; CGGA, Chinese Glioma Genome Atlas; CIC, Capicua Transcriptional Repressor; CNN, convolutional neural network; CPH, Cox proportional hazard; CRC, colorectal carcinoma; CRGs, cuproptosis-related genes; CRP, C-reactive protein; CTLA-4, cytotoxic T-lymphocyte antigen 4; CUB, Charité Universitätsmedizin- Berlin; DAE, deep autoencoders; DBN, deep belief networks; DCB, durable clinical outcome; DFCI, Dana-Farber Cancer Institute; DL, deep learning; DLS, deep learning score; dNLR, derived neutrophil-to-lymphocyte ratio; DNN, deep neural network; DTC, decision tree classifier; Enet, elastic network; eRNA, enhancer RNA; ETC, Extra Trees Classifier; Ex. Val, external validation; FAHZ, First Affiliated Hospital Zhengzhou; FCNA, fraction of copy number alteration; FFNN, feedforward neural network; FHFM, Flatiron Health and Foundation Medicine; GB, gradient boosting; GBC, gradient boosting classifier; GC, gastric cancer; GEO, gene expression omnibus; GLM, generalised linear models; GO, Gene Ontology; GPC, Gaussian process classifier; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; GUF, Goethe University-Frankfurt; Hb, haemoglobin; HED, human leukocyte antigen -I evolutionary divergence; HKEBHB, Helios Klinikum Emil von Behring Hospital Berlin; HLA LOH, loss of heterozygosity status in human leukocyte antigen —I; HNSCC, head and neck squamous cell carcinoma; HR, hazard ratio; ICI, immune checkpoint inhibitors; ICS score, immune cell signature score; IDH1, isocitrate dehydrogenase; iDMCs, immunophenotype-specific differentially methylated CpG sites; IO, immunotherapy; iPMS, immunophenotype-related methylation signature; IRLS, immune-related lncRNA signature; ITH, intratumoural heterogeneity; K-M, Kaplan—Meier; k-NN, k-nearest neighbours; LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis; LDH, lactate dehydrogenase; LGG, low-grade glioma; LightGBM, light gradient boosting machine; LMC, latent methylation components; lncRNAs, long non-coding RNAs; LR, logistic regression; MAGE-A3, melanoma-associated antigen 3; MARS, multivariate adaptive regression splines; MCC, Matthew's correlation coefficient; MeDeCom, reference-free computational framework that allows the decomposition of complex DNA methylomes into latent methylation components and their proportions in each sample; ML, machine learning; MLPS, machine learning-based prognostic signature; MLR, multivariate logistic regression; MNN, morphogenic neural networks; mRMR, minimum redundancy maximum relevance; MSC, blood microRNA signature classifier; MSI, microsatellite instability; MSKCC, Memorial Sloan—Kettering Cancer Center; NB, naïve Bayes; NCCJ, National Cancer Center Japan; NDB, no durable benefit; NDB-LS, long-term survival with no durable clinical benefit; NLR, neutrophil-to-lymphocyte ratio; NN, neural network; NPV, negative predictive value; NSCLC, non-small-cell lung cancer; ORR, objective response rate; OS, overall survival; PAC, proportion of ambiguous clustering; PAM, partition around medoids; PCA, principal component analysis; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; PFS, progression-free survival; plsRcox, partial least squares regression for Cox; plsRcox, partial least squares regression for Cox; PPV, positive predictive value; PS, performance status; PUMCH, Peking Union Medical College Hospital; QDA, quadratic discriminant analysis; rbf, radial basis function; RCC, renal cancer cell; RF, random forest; RF11, random forest classifier with 11 input features; RF16, random forest classifier with 16 input features; RFB, random forest and Boruta; RFS, relapse-free survival; RSAGs, robust survival-associated genes; RSF, random survival forest; SCNA, somatic copy number alteration; SelectKBest, selects the features according to the k highest score; Sen, sensitivity; SMC, Samsung Medical Center; SMS, somatic mutation signature; Spec, specificity; SuperPC, supervised principal components; SVM, support vector machine; SVM-RFE, support vector machine-recursive and feature elimination; T, training; TEX, T-cell exhaustion; TIDE, tumour immune dysfunction and exclusion; TIIClncRNA, tumour-infiltrating immune cell-associated lncRNA; TILBlncSig, lncRNA signature of TIL-Bs; TIL-Bs, tumour-infiltrating B lymphocytes; TILs, tumour-infiltrating lymphocytes; TISCH, Tumour Immune Single-cell Hub 2; TLS, tumour mutational burden-related LASSO score; TMUGH, Tianjin Medical University General Hospital; Tregs, T regulatory cells; TTF, time to treatment failure; UC, urothelial carcinoma; UHW, University Hospital Würzburg; V, Testing; VIMP, variable importance; WBC, white blood cell; WGCCES, West German Cancer Center Essen Heckeshorn; XGBDART, XGBoost Dropouts meet Multiple Additive Regression Trees; XGBoost, eXtreme Gradient Boosting; XHACSU, Xiangya Hospital Affiliated to Central South University; YCH, Yonsei Cancer Hospital.

https://doi.org/10.1016/j.annonc.2023.10.125

Table 3. Studies using radiomics features and artificial intelligence to predict immunotherapy response in cancers

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Dercle et al., 2020[78]	Retrospective analysis of prospective trials	Clinical trials (CheckMate017, Check-Mate063, CheckMate017)	NSCLC	188 (total) 92 (IO total) 72 (training) 20 (test) 50 (chemo total) 46 (TKI total)	Nivolumab	Delta-volume, delta-GLCM IMC1, delta-DWT1, and delta-sigmoid slope	Coarse selection (reproducibility analysis, redundancy analysis, and feature ranking) Fine selection (forward search, feature combination)	ML: classification; RF Survival analysis (statistics) CPH, Kaplan–Meier analysis; stratification between with and without high-risk nivolumab signature	Sensitive[a] versus non-sensitive to nivolumab AUC = 0.77; Sensitivity 0.80; Specificity 0.53; PFS $P < 10^{-4}$
He et al., 2020[79]	Retrospective	Institutional (SPH)	NSCLC	327 (total) 236 (training) 26 (test) 65 (test) 123 (IO validation)	Anti-PD-(L)1	1688 pre-defined radiomics features 1020 DL features TMB radiomics biomarker and ECOG PS	NA	DL (for creating a score TMBRB); Feature extraction and classification; Densenet with a 3D convolution kernel (3D densenet) Survival analysis (statistics) K-M curves for evaluating TMBRB risk stratification	Low- versus high-risk group: Best results achieved using TMBRB and ECOG PS OS: $P = 0.007$; PFS: $P = 0.003$
Khorrami et al., 2020[80]	Retrospective	Institutional (CCF, UPHS)	NSCLC	139 (total) 50 (training) 62 (validation1) 27 (ex. validation)	Anti-PD-(L)1	8 Del-RADx features	Intraclass correlation coefficient; WLCXfeature selection	ML Classification: LDA; DelRADx risk score Survival analysis (statistics): DelRADx risk score Kaplan–Meier, log-rank statistical tests	Responders versus non-responders Validation1 AUC 0.85 Validation2 AUC 0.81 Validation1 c-index 0.69 Validation2 C-index 0.68 OS $P = 00056$
Mu et al., 2021[81]	Retrospective/ prospective	Institution (SPH, MCC)	NSCLC	697 (total) 284 (training) 116 (testing) 85 (ex validation) 128 (IO validation) 49 (IO pros. validation) 35 (IO ex. validation)	Anti-PD-(L)1	Deep learned score Clinical features: ECOG, histology	NA	DL SResCNN to develop a deep learned score Survival analysis: (statistics) K-M method, stratification between low versus high deep learned score Cox proportional hazards model; and multivariable LR	DCB Retrospective validation AUC 0.70 $P < 0.001$, Prospective validation AUC 0.72 $P = 0.014$ Ex. Val AUC 0.70 $P = 0.040$ Retrospective val PFS $P < 0.001$ OS $P < 0.001$ Prospective val PFS $P < 0.001$ OS $P = 0.003$ Ex. val PFS $P = 0.038$ OS $P = 0.007$ Retrospective Val

Continued

Table 3. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
									DCR C-index 0.87 PFS C-index 0.73 OS C-index 0.0.77 Prospective Val DCR C-index 0.82 PFS C-index 0.74 OS C-index 0.0.70
Tian et al., 2021[82]	Retrospective	Institutional (WCHSU)	NSCLC	939 (total for sig. creation) 750 (training) 93 (test) 96 (validation) 94 (IO validation set)	Pembrolizumab ± chemotherapy	Deep learning features, radiomics features (shape, texture, first-order), clinical characteristics	Mann−Whitney *U* test; DL network	DL: classification; network: deep learning feature extraction module (densenet121), a conventional radiomic feature extraction, a fully connected classification layer; creation of PD-L1 ES signature Survival analysis (statistics): Kaplan−Meier curves and log-rank test, stratification using PD-L1ES score	PFS C-index 0.66, 95% CI 0.48∼0.83 HR 2.57, 95% CI 1.22∼5.44; $P = 0.010$
Tunali et al., 2019[83]	Retrospective analysis of prospective trials	Clinical trials at MCC (not specified)	NSCLC	228 (total)	Anti-PD-(L)1	Radiomics features: Radial gradient border SD-2D, 3D Laws E5L5E5, border 3D Laws E5E5L5 and border quartile coefficient of dispersion, border NGTDM strength Clinical features lines of therapy, hepatic and bone metastasis, NLR, RMH prognostic score	LR	ML CART; Survival analysis: K-M analysis	TTP AUC 0.804; Spec. 83.46% Sen. 63.43%; Acc. 73.41% HPD AUC 0.865, Spec. 74.02%; Sen. 90.55%; Acc. 82.28% PFS AUROC = 0.717
Tunali et al., 2021[84]	Retrospective analysis of prospective trials	Institutional (MCC, JAHVHT)	NSCLC	332 (total) 180 (training) 90 (test) 62 (validation)	Anti-PD-(L)1 ± anti-CTLA-4	Clinical features: serum albumin, metastatic sites Radiomic features: GLCM inverse difference Avg 3D RLN normalised, GLCM inverse variance, Avg 3D SRE, Avg 3D RP, GLSZM Zone percentage, Avg 3D LRE, Avg 3D RLV, peritumoral quartile coefficient of dispersion, peritumoral coefficient of variance	Concordance correlation coefficient	ML: a survival CART Survival analysis: K-M, Cox regression, log-rank test	OS; time-dependent AUC: 6 months = 0.729, 95% CI 65.2-83.8; 12 months = 0.773, 95% CI 71.7-83.7; 24 months = 0.709, 95% CI 72.5-87.5; 36 months = 0.868, 95% CI 77.0-93.4
Vaidya et al., 2020[85]	Retrospective	Institutional (CCO)	NSCLC	109 (total) 30 (training) 79 (testing)	Anti-PD-(L)1	One peritumoral texture and two QVT features (mean curvature of branches surrounding the tumour,	Unsupervised clustering on the radiomics features, MRMR	ML Classifiers: RF, linear discriminant analysis, diagonal linear discriminant analysis,	Responders and non-responders versus hyperprogression Random forest: AUC 0.96; Acc. 0.83; Sen.

Continued

https://doi.org/10.1016/j.annonc.2023.10.125

Table 3. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
						curvature values associated with vasculature)		quadratic discriminant analysis, SVM. Survival analysis (statistics): K-M; stratification between classifier prediction	1.0; Spec 0.8 OS HR = 2.66, 95% CI 1.27-5.55; $P = 0.009$
Deng et al., 2022[86]	Retrospective	Institutional (SPH, GPPH, WCH, FAHUSTC, LCHHI)	NSCLC	699 (total) 386 (training) 92 (test) 92 (ex. validation) 129 (IO validation)	ICIs		NA	DL classifier: ESBP; Survival analysis: (statistics) Kaplan −Meier, log-rank test, stratification using ESBP score high versus low score	PFS HR = 0.36, 95% CI 0.19-0.68, $P <$ 0.0001) Survival benefit HR = 0.33, 95% CI 0.18-0.55, $P < 0.0001$
Gong et al., 2022[87]	Retrospective	Institutional (SPH, FUSCC)	NSCLC	224 (total) 93 (training) 68 (test1) 63 (test2)	Anti-PD-(L)1	Radiomic features: original shape_ maximum 2D diameter row, original_girlm_ LongRunHighGrayLevelEmphasis, LoG-glszm_SizeZoneNonUniformity, wavelet-LHH_glszm_Large AreaLowGrayLevel Emphasis, wavelet-HLL_glcm_MCC, wavelet-HHH_firstorder_Skewness, wavelet-LLL glcm _MCC; Delta-radiomic features: wavelet-LLH first-order range, wavelet-LLH firstorder_RootMeanSquared, wavelet-LLH glszm SmallAreaEmphasis, wavelet-HHH firstorder_TotalEnergy	Feature ranking with RFE	ML Classification: SVM Survival analysis: K-M	Responders versus non-responders Test1: AUC 0.82 Acc. 76.47% Sen. 66.67% Spec. 79.25% F1 0.56 Test2 AUC 0.87 Acc. 80.25 % Sen. 79.17% Spec.82.05% F1 0.76 Survival analysis Test 1 PFS, HR = 6.10, 95% CI 2.12-17.56, $P <$ 0.001 OS, HR = 3.17, 95% CI 1.19-8.41, $P <$ 0 .05 Test 2 PFS, HR = 4.55, 95% CI 1.89-10.92, $P <$ 0.001 OS, HR = 2.95, 95% CI 1.11-7.84, $P < 0.05$
He et al., 2022[88]	Retrospective	Institutional (SPH)	NSCLC	236 (total) 188 (train) 48 (test)	Anti-PD-(L)1	Deep learning features	NA	DL: survival network with two modules: convolutional module and classification module; dual-task network for PDS and PRS; (for OS and PFS risk factors computation) Survival analysis: K-M	Stratification using OS: C-index 0.75 HR = 4.54, 95% CI 1.21-16.94, log-rank $P = 0.014$ PFS C-index 0.7 HR = 6.64, 95% CI

Continued

Table 3. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
								curve and log-rank test; stratification using risk score from DL	2.89-15.29, log-rank $P < 0.001$
Ren et al., 2022[89]	Retrospective	Institutional (UHTMC)	NSCLC	157	Nivolumab, pembrolizumab	Radiomics features: GLCM, GLSZM, GLRLM, NGTDM, GLDM Deep learning features	CHSQ, RELF, MIM, FSCR, MIFS, GINI, ICAP, JMI, CIFE, CMIM, DISR, MRMR, TSCR	ML: LR, k-NN, QDA, SVM with linear and radial basis function kernels, XGBoost, multilayer perceptrons, Gaussian processes, decision trees, nBayes, RF, AdaBoost	InceptionV3_RELF_ k-NN (best model): AUC 0.96; Acc. 95.24%; Sen. 95%; Spec. 95.5%; Pre. 1.67%; F1 95.30%
Dercle et al., 2022[90]	Retrospective analysis of prospective trials	Clinical trials (Keynote-002, Keynote-006)	Melanoma	575 (total) 288 (training) 287 (test)	Pembrolizumab	Volumetric growth (absolute tumour volume difference), tumour volume, quantitative representation of tumour spatial heterogeneity, quantitative representation of tumour edge phenotype	PCA	ML: classification, RF Survival analysis: K-M	OS AUC = 0.92 (95%CI 0.89-0.95)
Basler et al., 2020[91]	Retrospective	Institutional (UHZ)	Melanoma	112 (total: 10-fold cross-validation)	Anti-PD-1 ± anti-CTLA-4	Blood radiomics features: LDH level at TP1 and the relative change of CT coarseness between TP1 and TP0	Pearson correlation coefficient Univariate supervised feature selection (*F*-test) LR model regularised with elastic net	ML: classification: LR and regularised with L2 penalty Survival analysis (statistics): K-M, Mann—Whitney *U* test	PP versus TDP; AUC 0.82; Sen. 0.81; Spec 0.73 PPV 0.54; NPV 0.91
Brendlin et al., 2021[92]	Retrospective	Database from institution ISC	Melanoma	140 (total) 70 (train) 70 (test)	Anti-PD-1 anti-CTLA-4, Anti-PD-1 + anti-CTLA-4	DECT: Mean iodine, iodine concentration total, mean mixed, min. brightness, textural coarseness, mean absolute voxel intensity deviation	Univariate statistic with determination coefficient (R^2), mutual information, Bonferroni correction for multiple comparisons, minimum redundancy Maximum relevance algorithm, stepwise forward selection	ML: classification: RF	Responders versus non-responders Best model trained on DECT: AUC 0.85 Sen. 74.63% Spec. 75.90% Acc. 75.26%
Peisen et al., 2022[93]	Retrospective	CMMR	Melanoma	262 (total; fivefold cross-validation)	Anti-PD-1/CTLA-4 monotherapy/ combination therapy	Clinical and radiomic features Addition of radiomic feature did not improve the prediction	FCBF	ML: classification, RF	PFS (3 months): AUC 0.641 OS (6 months): AUC 0.664 OS (12 months): AUC 0.600
George et al., 2022[94]	Retrospective—*post hoc* analysis	Clinical trial (NCT02336165)	GBM	113 (total)	Durvalumab	Clinical features Radiomic features: Spherical disproportion, GLCM autocorrelation, GLSZM SZHGE, Histogram variance, GLCM sum average, GLCM	Spearman correlation	ML: Survival analysis; RSF	OS C-index = 0.692-0.750 PFS C-index = 0.680-0.715

Continued

https://doi.org/10.1016/j.annonc.2023.10.125

Table 3. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
						correlation, maximum axial diameter, GLSZM GLN, GLSZM SZLGE, maximum axial diameter, assign (treatment regimen), histogram root mean squared, histogram variance, histogram kurtosis, histogram percentile 50, GLSZM LGZE, histogram SD, GLRLM LRLGE			
Trebeschi et al., 2019[95]	Retrospective	Institutional (NKI, MCC)	Melanoma, NSCLC	203 (total) 133 (train) 70 (test)	Anti-PD-1	Radiomics features	Feature selectors: unsupervised resulting from PCA, supervised resulting from WFS	ML: Genetic algorithms, RF, LR Survival analysis (statistics): K-M for OS	LR: AUC = 0.62 ± 0.01 RF: AUC = 0.64 ± 0.02
Sun et al., 2019[96]	Retrospective	Institutional (GRI), clinical trial (MOSCATO), TCGA, TCIA	Multiple cancer types	385 (total) 135 (train) 119 (validation) 137 (ex IO validation)	Anti-PD-(L)1	Radiomic features (texture, histogram, volume), peak Kilovoltage, VOI location	Linear elastic net model	ML: Elastic net regularised regression method for creation of radiomics-based CD8 cell score Survival analysis: K-M Method, stratification using radiomics-based score on high and low, Wilcoxon and log-rank Cox's proportional hazards model	AUC of the radiomic signature = 0.67, 95% CI 0.57-0.77 Responders versus non-responders $P = 0.013$ OS: HR 0·58, 95% CI 0.39-0.87, $P = 0.0081$
Ligero et al., 2021[97]	Retrospective	Database from institution and clinical trials	Multiple cancer types, breast, cervix, gastrointestinal	178 (total) 85 (training) 46 (test1) 47 (test2)	Anti-PD-(L)1	Clinical features: baseline albumin level, lymphocyte count; Radiomic features: histogram, shape Local-regional texture features	Elastic net model	ML: Regression model Survival analysis: Kaplan—Meier	AUC = 0.74, 95% CI 0.63-0.84; Sensitivity = 77%; Specificity = 59%

3D, three-dimensional; Acc, accuracy; AdaBoost, adaptive boosting; AUC, area under the curve; AUC-ROC, area under the receiver operating characteristic curve; CART, classification and regression tree; CCF, Cleveland Clinic Foundation; CCO, Cleveland Clinic Ohio; CHSQ, chi-square score; CI, confidence interval; CIFE, conditional infomax feature extraction; CMIM, conditional mutual information maximisation; CMMR, Central Malignant Melanoma Registry; CPH, Cox proportional hazard; CT, computed tomography; CTLA-4, cytotoxic T-lymphocyte antigen 4; DCB, durable clinical benefit; DCR, disease control rates; DECT, dual-energy specific radiomic features; DelRADx, delta in the radiomic texture; DISR, double input symmetric relevance; DL, deep learning; DWT1, discrete wavelet transform 1; ECOG PS, Eastern Cooperative Oncology Group performance status; ESBP, EfficientNetV2-based survival benefit prognosis; Ex. Val, external validation; Exp, exploratory analysis cohort; FAHUSTC, The First Affiliated Hospital of University of Science and Technology of China; FCBF, fast correlation-based filter; FSCR, Fisher score; FUSCC, Fudan University Shanghai Cancer Center; GBM, glioblastoma; GINI, gini index; GLCM, gray-level co-occurrence matrix; GLCM IMC1, gray-level co-occurrence matrix informal measure of correlation1; GLDM, gray-level dependence matrix; GLRLM, gray-level run-length matrix; GLSZM, gray-level size zone matrix; GPPH, Guangdong Provincial People's Hospital; GRI, Gustave Roussy Institute; HPD, hyperprogressive disease; HR, hazard ratio; ICAP, interaction capping; ICI, immune checkpoint inhibitor; IMC, immunochemistry; IO, immunotherapy; ISC, Imaging Science Center, Department of Diagnostic and Interventional Radiology, Universitätsklinikum Tübingen, Tubingen, Germany; IT, immunotherapy cohort; JAHVHT, James A. Haley Veterans' Hospital—Tampa; JMI, joint mutual information; K-M, Kaplan—Meier; k-NN, k-nearest neighbours; LCHHI, Liaoning Cancer Hospital and Institute; LDA, linear discriminant analysis; LDH, lactate dehydrogenase; LGZE, low gray-level zone emphasis; LR, logistic regression; LRLGE, long run low gray-level run emphasis; MCC, Moffitt Cancer Center; MIFS, mutual information feature selection; MIM, mutual information maximization; ML, machine learning; MRMR, minimum redundancy maximum relevance; NA, not applicable; NGTDM, neighbourhood gray-tone difference matrix; NGTDM, neighbourhood gray-tone difference matrix; NKI, Netherlands Cancer Institute; NLR, neutrophil-to-lymphocyte ratio; NPV, negative predictive value; NSCLC, non-small-cell lung cancer; OS, overall survival; PCA, principal component analysis; PD-(L)1, programmed death-ligand 1 (PD-L1)/programmed death protein 1 (PD-1) *PD-L1ES*- PD-L1 expression signature; PDS, progressive disease score; PET, positron emission tomography; PFS, progression-free survival; PP, pseudoprogression; PPV, positive predictive value; Pre, precision; PRS, partial response score; QDA, quadratic discriminant analysis; QVT, quantitative vessel tortuosity; RELF, ReliefF; RF, random forest; RFE, recursive feature elimination; *RLV*, run-length variance; RMH, Royal Marsden Hospital; RP, run percentage; RSF, random survival forest; Sen, sensitivity; Spec, specificity; SPH, Shanghai Pulmonary Hospital; SRE, short runs emphasis; SResCNN, small residual convolutional network; SVM, support vector machine; SZHGE, short-zone high gray-level emphasis; SZLGE, short-zone low gray-level emphasis; T, training cohort; TCGA, The Cancer Genome Atlas; TCIA, The Cancer Imaging Archive; Te, test cohort; TMB, tumour mutational burden; TMBRB, tumour mutational burden radiomic biomarker; TP0, baseline; TP1, timepoints 1-3 months; TPD, true progressive disease; TSCR, test score; TTP, time to progression; UHTMC, Union Hospital—Tongji Medical College; UHZ, University Hospital Zurich; UPHS, University of Pennsylvania Health System; V, validation cohort; VOI, volumes of interest; WCH, West China Hospital; WCHSU, West China Hospital of Sichuan University; WFS, wrapper feature selection; WLCX, Wilcoxon rank sum; XGBoost, eXtreme Gradient Boosting.

[a]Any early tumour changes reflecting a possible improvement of the response.

classifying responders and non-responders (AUC 0.69-0.79) across three different external datasets. Only 14% ($n = 5$) of included studies used DL techniques. Fang et al.[37] developed a deep patient graph convolutional network (DeePaN), integrating 100 electronic health records (EHR) and genomic data features to divide patients into five subgroups that showed significant OS differences (P value < 0.0001). Codes are publicly available for 35% ($n = 13$) of the included studies.

Several of the reviewed studies have used AI tools showing a better performance compared to standard ICI biomarkers. Results confirmed TMB, KRAS, TP53, STK11, and MSI as the most relevant on predictive AI algorithms when integrated with additional biomarkers [e.g. clinical characteristics, neutrophil-to-lymphocyte ratio (NLR), lactate dehydrogenase (LDH), albumin, disease burden, PD-L1]. Again, a TMB related LASSO score (TLS) was used to stratify urothelial carcinoma patients, between responders and non responders to ICI.[39] Finally, new candidate biomarkers have been discovered with such AI tools e.g. loss of heterozygosity status in human leukocyte antigen (HLA LOH) and genomic intra-tumour heterogeneity (ITH).

Transcriptomics. Studies that used transcriptomics data manly reported gene signature-based models which outperform conventional biomarkers such as PD-L1 in different cancer types.[43,49,52]

Moreover, the use of AI tools has allowed for the discovery of previously unknown biomarkers. For instance, Charoentong et al.[44]used ML to identify determinants of tumour immunogenicity, resulting in the development of the immunophenoscore. Zheng et al.[55] developed a predictive model incorporating biomarkers to achieve an AUC of 93% in clear cell renal cancer carcinoma treatment, indicating that AI algorithms can identify more precise and effective biomarkers. At length, new RNA-based biosignatures have been generated with AI transcriptomics such as the signature of cancer stemness, cuproptosis, AGR (angiogenesis), HLA presentation, T-cell exhaustion, and chemokine signalling.

Epigenomics. AI has shown significant potential in the discovery of epigenomics biomarkers. Firstly, many long non-coding (Inc) RNA-based signatures have been discovered: a tumour-infiltrating immune cell-associated lncRNA (TIIClncRNA), immune-related lncRNA signature (IRLS), tumour-infiltrating B lymphocytes lnc signature (TILBlncSig), and ML models in low-grade gliomas, colorectal carcinoma and different type of cancers, respectively, for predicting ICI efficacy or as an independent risk factor.

Additionally, AI-driven approaches have been applied to the analysis of DNA methylation data. Filipski et al.[66] revealed that latent methylation components-based signatures were predictive for ICI in metastatic melanoma patients. Xu et al.[67] demonstrated the potential of DNA methylation profiles in predicting ICI response across cancers, using a high-performing SVM model. Pan et al.[70] identified five CpG sites in an immunophenotype-related methylation signature, which provided immune heterogeneity information and potential clinical treatment guidance for lung adenocarcinoma.

Radiomics

Our inclusion criteria identified 20 studies using radiomics features (Table 3). Since many papers used a small number of patients (<100 patients), in Supplementary Information 3, Table S3, available at https://doi.org/10.1016/j.annonc.2023.10.125, we included 11 studies that were initially excluded using that criterion. Most of the studies (60%, $n = 12$) are done on NSCLC patients, followed by studies done on melanoma (20%, $n = 4$). Codes are publicly available only for three studies (15%). Broadly, two typical workflows are present in the selected studies. The first includes manual or semi-automated image segmentation followed by feature extraction that is not part of the learning process, feature selection, and finally ML prediction (75% of the studies, $n = 15$). The second approach includes the use of DL models (25% of the studies, $n = 5$), mainly CNN models, where segmentation can be incorporated into the DL architecture, features are not manually defined and selected, and the model learns from raw data.

One of the largest and most recent studies that used the first, ML-based approach was an analysis on 575 melanoma patients treated with ICI in the prospective KEYNOTE-002 and KEYNOTE-006 trials, conducted by Dercle et al.[90]. Using an RF algorithm, a radiomic signature combining four imaging features (two related to tumour size and two reflecting changes in tumour imaging phenotype) was identified and showed a higher performance in estimating OS compared to the standard method (RECIST criteria) (AUC 0.92 versus 0.80). Among studies using the second, DL-based approach, a study conducted by Tian et al.[82] where DL CNN for extracting DL features and a fully connected network that combines DL features, pre-defined radiomic features, and clinical features was used to create a PD-L1 expression signature. This signature was further validated to predict PFS using Kaplan—Meier group stratification on high and low risk ($P = 0.01$).

Related to meta-biomarker discovery, the importance of different features is reflected directly within the developed algorithm and therefore it is difficult to always draw conclusions on a single biomarker value and on one which is giving weight to the ICI prediction. Some notable biomarkers discovered include predictive signatures for NSCLC treatment sensitivity to nivolumab, docetaxel, and gefitinib[78]; a TMB radiomic-based biomarker[79]; and a deeply learned score (DLS) for predicting programmed death-ligand 1 (PD-L1) expression.[81] Moreover, studies have identified potential radiomics markers for rapid PD,[83,91] survival outcomes prediction,[84,90] and immunotherapy response prediction in NSCLC and melanoma patients.[89,92] Furthermore, radiomic signatures have been developed to predict tumour immune phenotype and clinical outcomes for patients treated with ICI.[96,97] Interestingly, some papers (10%) have also been focused on peritumoral texture[84,85] by exploring the TME. Finally, interestingly, some radiomic features

Table 4. Studies using digital pathology features and artificial intelligence to predict immunotherapy response

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Harder et al., 2019[98]	Retrospective	Clinical trial (MISIPI study), institutional (INT-Pascale)	Melanoma	31 (total; cross-validation)	Ipilimumab	Pathomics: CD3, CD8, and FoxP3	Monte Carlo cross-validation	ML: Cell-free segmentation and cell-free classification—learning-based annotation-free approach and ANOVA for model comparison DL: Region classification for accurate differentiation between melanin and IHC marker-positive immune cells —CNN	Best model: decision tree model based on relative densities of CD8+ TILs in the intratumoural infiltration region. CNN accuracy: 98.2%
Park et al., 2022[99]	Retrospective	TCGA; institutional (SMC, SNUBH)	NSCLC	3017 (total) 2389 (AI algorithm development) 518 (patients evaluated for survival) 110 (external validation cohort)	Immune checkpoint inhibitors	Pathomics: IP assessed by spatial distribution of TILs; AI-powered spatial TIL analyser (Lunit SCOPE IO)	NA	ML: classification; Lunit SCOPE IO Survival analysis (statistics): K-M	PFS (months; Kaplan −Meier): 4.1 with inflamed IP, 2.2 with immune-excluded IP, 2.4 with immune-desert IP $P = 9.6 \times 10^{-5}$ and 4.1×10^{-4} OS (months; Kaplan −Meier): 24.8 with inflamed IP, 14.0 with immune-excluded IP, and 10.6 with immune-desert IP. $P = 0.023$ and 0.002
Baxi et al., 2022[100]	Retrospective	Clinical trials (CheckMate 275, CheckMate 026, CheckMate 057, CheckMate238, CheckMate 141, CheckMate 067)	Urothelial carcinoma, melanoma, NSCLC, HNSCC	4248 (total slides) 1960 (training set slides) 2288 (test set slides)	Nivolumab; Nivolumab plus ipilimumab	Pathomics: PD-L1 expression	Clinical (based on pathologists' expertise)	DL: CNN	High correlation in PD-L1 assessment between DL model and pathologists (r-score ranging from 0.73 to 0.85); Similar prediction of ORR with manual and AI-powered PD-L1 evaluation (respectively, AUC = 0.596 and AUC = 0.602)
Althammer et al., 2019[101]	Retrospective	Clinical trial (Study 1108/ NCT0169356)	NSCLC	362 (163 treated with immunotherapy and 199 not treated with immunotherapy)	Durvalumab	Pathomics: PD-L1 and CD8 expression	Clinical (based on pathologists' expertise)	DL (not specified)	OS (by multivariate Cox regression analysis) was significantly longer for signature-positive patients (PD-L1 and CD8+) compared with signature-negative patients (21.0 versus 7.8 months; $P = 0.00002$)
Johannet et al., 2021[102]	Retrospective	Institutional (NYUPCCC, VUICC)	Melanoma	151 (total) 121 (training set) 30 (independent validation set)	Anti-PD-1 and/or anti-CTLA-4	Multivariable classifier including DL and clinical data; cell nuclei	Backward stepwise selection	DL: Segmentation and response classification: DCNN	The multivariate model was able to predict ORR with AUC = 0.8 and to categorise patients in high

Continued

Table 4. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
						important for the prediction		Inception v3 as foundation architecture	versus low risk of progression ($P = 0.003$)
Hu et al., 2021[103]	Retrospective	TCGA-SKCM, institutional (PUCH, GPCH)	Melanoma (training and validation) and lung cancer (only validation)	299 (total) 190 (training set) 109 (ex. validation set)	Anti-PD-1	Prediction of response to anti-PD-1	PCA	DL: Xception (neural network with ImageNet pre-trained parameters)	Prediction of response to anti-PD-1 (melanoma): AUC = 0.778 Prediction of response to anti-PD-1 (lung cancer): AUC = 0.645
Chen et al., 2022[104]	Retrospective	Institutional (PUCH); clinical trials (NCT03472365, NCT03713905).	Gastric cancer	80 (total enrolled patients) 59 (patients treated with immunotherapy) 44 (training cohort—treated with immunotherapy) 15 (validation cohort—treated with immunotherapy)	Anti-PD-1 or anti-PD-L1	Pathomics: Density and spatial patterns of tumour-infiltrating immune cells	Threefold cross-validation and multiple repetitions of the whole prediction process (due to some classifiers having random starting points)	Image analysis system (inForm) ML: classification: ETC, ABC, GBC, and MLP	Higher levels of specific tumour-infiltrating T-cell subsets are associated with shorter OS in the global population. Univariate Cox proportional hazard regression model was used for calculating HR of each indicator in survival analysis
Choi et al., 2022[105]	Retrospective	Multiple institutions (SMC, Seoul National University Bundang Hospital)	NSCLC	802 (development) 479 (ex. Validation) Among them 430 used as IO test	ICI	Pathomics: TPS	NA	DL: CNN (faster R-CNN) for TPS score creation Survival analysis (statistics) for stratification based on TPS score	Two groups: PFS HR = 0.62, $P = 0.002$ OS HR = 0.61 $P = 0.013$

ABC, AdaBoost classifier; AI, artificial Intelligence; ANOVA, analysis of variance; AUC, area under the curve; CD3, cluster of differentiation 3; CD8, cluster of differentiation 8; CNN, convolutional neural networks; CTLA-4, cytotoxic T-lymphocyte antigen 4; DCNNs, deep convolutional neural networks; ETC, extra tree classifier; FOXP3, forkhead box P3; GBC, gradient boosting classifier; GPCH, Guangdong Province Cancer Hospital; HNSCC, head and neck squamous cell carcinoma; HR, hazard ratio; ICI, immune checkpoint inhibitor; IHC, immunohistochemistry; INT, Istituto Nazionale Tumori; IO, immunotherapy; IP, immunophenotype; K-M, Kaplan—Meier; L, deep learning; MISIPI, Melanoma ImmunoScore evaluation in patients treated with ipilimumab; ML, machine learning; MLP, multilayer perceptron; NSCLC, non-small-cell lung cancer; NYUPCCC, New York University Perlmutter Comprehensive Cancer Center; ORR, objective response rate; OS, overall survival; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; PFS, progression-free survival; PUCH, Peking University Cancer Hospital; SKCM, skin cutaneous melanoma; SMC, Samsung Medical Center; SMC, Samsung Medical Center; SNUBH, Seoul National University Bundang Hospital; TCGA, The Cancer Genome Atlas; TILs, tumour-infiltrating lymphocyte; TPS, tumour proportion score; VUICC, Vanderbilt University Ingram Cancer Center.

https://doi.org/10.1016/j.annonc.2023.10.125

Table 5. Studies using real-world features and artificial intelligence to predict immunotherapy response in cancers

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
Real-world data									
Benzekry et al., 2021[108]	Retrospective	Institutional (CRCM, AP-HM, GRI)	NSCLC	298 (total; cross-validation)	anti-PD-L1, anti-PD-1, anti-CTLA-4	PS, Hb, NLR	RF	ML, classification Multiple classifiers, best-performing: RF	DCR: Acc. 0.68 AUC-ROC 0.74 PPV 0.70 NPV 0.68 Sensitivity 0.58 Specificity 0.78
Hellwig et al., 2021[109]	Retrospective analysis on prospective cohort	CheckRad-CD8 trial	HNSCC	22 (total; cross-validation)	Tremelimumab; durvalumab	DCE	Sequential backward selection based on parallel random forests	ML, classification RF	Response versus non-response: sensitivity 78.7%; specificity 78.6%; AUC 0.866
Weber et al., 2017[110]	Retrospective analysis on prospective data	Clinical trial (NCT01176461, NCT01176461) + institutional (YSPORE-SC, INT—Pascale, MSCC)	Melanoma	289 (total) 119 (training set) 170 (total test sets from five different sets)	Nivolumab, pembrolizumab, ipilimumab	Serum protein signature —PSEA	k-NN	ML for creating signature Survival analysis: k-NN combined via LR	Results on test set: OS: $P < 0.001$ PFS: $P = 0.236$
Lui et al., 2022[111]	Retrospective	Institutional (CDARS)	HCC	436 (total) 316 (training) 79 (test) 43 (ex. validation)	Ipilimumab, nivolumab, pembrolizumab	Shapley method: AFP, bilirubin, and alkaline phosphatase	Forward and backward stepwise selection of the LR	ML, classification; LR, LASSO, XGB, RF, GBM, NN Survival analysis; XGB time-to-event model	OS 1 year; best-performing RF Internal validation; AUC-ROC 0.92; sensitivity 0.84; specificity 0.85; PPV 0.88; NPV 0.80; PLR 5.74; NLR 0.18; External validation cohort; RF AUC-ROC 0.91
Arbour et al., 2021[112]	Retrospective	Institutional (MSK-MGH)	NSCLC	543 (total) 361 (training set)) 92 (test set) 97 external validation set)	Anti-PD-(L)1	Scan reports for RECIST reads	Not reported	Fully connected DL; classification	Test set: BOR (three classes): Sen. class 1 = 84% Spec. class 1 = 96% Sen. class 2 = 80% Spec. class 2 = 88% Sen. class 3 = 89% Spec. class 3 = 93% PFS (two classes): Acc. 85% Progression date: Acc. 73% 2 months progression: Acc. 80% External set: BOR (three classes): Sen. class 1 = 69% Spec. class 1 = 94% Sen. class 2 = 43% Spec. class 2 = 72% Sen. class 3 = 70% Spec. class 3 = 75% PFS (two classes): Acc. 82%

Continued

Table 5. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
									Progression date: Acc. 59% 2 months progression: Acc. 82%
Wu et al., 2022[113]	Retrospective	Clinical trials (NCT0184641, NCT01903993, NCT02031458, NCT02008227, NCT02951767, NCT02108652, NCT0230280, NCT01984242)	NSCLC, bladder transitional cell carcinoma, RCC	2538 (total) 1776 (Training) 762 (test set)	Atezolizumab	CRP, PD-L1, cancer type, Liver metastasis, dNLR, alkaline phosphatase, albumin, Hb, WBC, number of metastasis sites, pulse rate, PS	LASSO	ML; Classification; XGB, RF, LR, LASSO, SVM and k-NN	Death risk (Y/N); Best-performing RF; Test set: AUC-ROC 0.79 Acc. 0.72 Sen. 0.78 Spec. 0.65 PPV 0.73 NPV 0.71 Kappa 0.44 F1 0.76 Brier 0.27
Li et al., 2022[114]	Retrospective	TCGA, GEO, Imvigor210	NSCLC	428 (total) 117 (training), 78 (test) 116 (ex. validation)	Anti-PD-1	CD8 + T cells and macrophages	Univariate COX Wilcoxon test LASSO	COX LASSO for creating signature PPS used in ML models ML: Classification; SVM, RF, XGB, Adaboost Survival analysis DL: Classification; NN	Test/external validation set: Adaboost: Acc. Cluster 1 0.96/0.84 Acc. Cluster 2 0.96/0.8 Acc. Cluster 3 1/0.96 AUC Cluster 1 0.96/0.75 AUC Cluster 2 0.94/0.79 AUC Cluster 3 1/0.9 F1 Cluster 1 0.96/0.90 F1 Cluster 2 0.93/0.76 F1 Cluster 3 1/0.97 Pre. Cluster 1 0.92/0.89 Pre. Cluster 2 1/0.82 Pre. Cluster 3 1/0.94 Rec. Cluster 1 1/0.91 Rec. Cluster 2 0.87/0.7 Rec. Cluster 3 1/1
Liang et al., 2022[115]	Retrospective	Clinical trials (NCT0177200, NCT0251739, NCT02496208, NCT00001506, NCT02471352)	Melanoma	167 (total) 88 (training) 40 (test set) 39 (external validation 1) 27 (external validation 2)	IO	Gut microbiome	Multivariate selbal analysis	ML: classification; LASSO, RT, RF, NN, SVM, SL	Responder versus non-responder Test set SVM: AUC 0.73 Ex. Val 1: AUC 0.67 Ex Val 2: AUC 0.72 Ex val combined: 0.63
Muller et al., 2020[116]	Retrospective	Institutional (NKI, VUMc, EMC), clinical trial (NCT00989690)	NSCLC	289 (total) 116 (training) 98 (ex. validation 1) 75 (ex. validation 2)	Nivolumab	Serum-derived protein	Deep matrix-assisted laser desorption/ ionization (MALDI), GSEA	ML and DL from: diagnostic Cortex™[a] data analysis platform: Multiple classifiers	Stratification between: Resistant versus non-resistant (OS) Ex. val 1 $P = 0.037$ Ex. val 2 $P = 0.007$ Sensitive versus non-sensitive (OS) Ex. val 1 $P = 0.038$ Ex. val 2 $P = 0.179$
Iivanainen et al., 2022[117]	Retrospective analysis on	KISS trial	NSCLC, melanoma, genitourinary	31 (total) (cross-validation)	Anti-PD-(L)1	Patient-reported data	Clinical	ML Classification XGBoost	ORR: Acc. 75%; AUC 0.71; F1

Continued

https://doi.org/10.1016/j.annonc.2023.10.125

Table 5. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
	prospective data		cancers, head and neck cancer						score 0.58; MCC 0.4; Sen. 0.58; Spec. 0.82
Wei et al., 2022[118]	Retrospective	Institutional (ATHZU)	NSCLC, RCC, EESC, GAC, other	120 (total with IO) 91 (training) 29 (test) 40 (without IO)	Camrelizumab Nivolumab Pembrolizumab Sintilimab Others ICI	PBIMs	ASR, LASSO regression Univariate Cox regression	ML classification SVM-RFE, RFB, and elastic net models. Survival analysis	NDB versus NDB_LS& DCB: Test set: SVM-RFE AUC-ROC 0.73 Acc. ∼0.75 Kappa ∼ 0.3
Sun et al., 2021[119]	Retrospective	TCGA, GEO, IMvigor210	Lung adenocarcinoma	1328 (total) 500 (training) 348 (IO test)	Atezolizumab	FRRS (top three genes CISD1, FANCD2, and SLC3A2)	Random survival forest and PCA	ML: Score creation: RSF and PCA Survival analysis (statistics) K-M	IO test set AUC 2-year 0.686 AUC 3-year 0.699 AUC 5-year 0.694
Abuhelwa et al., 2021[120]	Retrospective	Clinical trials (IMvigor210, IMvigor211)	Urothelial cancer	896 (total) 429 (training) 467 (ex. validation)	Atezolizumab	More clinical-pathological factors (curated list). Curated top features: CRP, alkaline phosphatase, NLR, LDH, tumour sites	Clinical	ML: Survival analysis; GBM, RF-SRC, Cox-boosted, GLM	OS (GBM): Ex. Val C-index = 0.71 PFS (all): Ex. Val C-index = 0.62
Gupta et al., 2021[121]	Retrospective analysis on prospective data	Clinical trial (CheckMate 025), IMDC	RCC	2955 (total) 803 (training) (cross-validation) 2152 (ex. validation)	Nivolumab	More clinical-pathological factors All group OS variables: MSKCC risk group, IMDC/Heng risk score, Karnofsky PS	Bayesian networks	ML: Classifiers BNM individual-level predictor; LASSO, SVM, and RF	10-fold cross-validation (BNM/RF): OS 12 months: AUC 0.74/0.77 AE 3 years: AUC 0.71/0.72 TRAE 2 years: AUC 0.72/0.75 Ex. val (BNM) OS 12 months: AUC 0.71
Sidhom et al., 2022[122]	Retrospective	Clinical trial (CheckMate-038)	Melanoma	43 (total) (cross-validation) 11 (ex. validation1) 19 (ex. validation1)	Anti-PD-1 or anti-PD-1 plus anti-CTLA-4	Tumour-specific TCRs	TCR featurisation; DL model	DL; Classification; Deep TCR's repertoire classifier Survival analysis (statistics) stratification in low and high likelihood of response groups using model prediction	Responder versus non-responder: Cross-val AUC 0.86 Ex. Val 1 AUC 0.82 Ex. Val 2 AUC 0.61 Ex. Val combined AUC 0.71 Low versus high likelihood of response (training) PFS $P = 0.005$
Rounis et al., 2021[123]	Prospective	Institutional (UGHH)	NSCLC	66 (total) (10-fold cross-validation)	Anti-PD-1/PD-L1	More clinicopathological information. Best: prolonged antibiotics administration	LASSO	ML Classification; multiple classifiers; best-performing SVM Applied using: JADBio[b]	Partial response or stable disease versus progression disease AUC 0.806
Madonna et al., 2021[124]	Retrospective	Institutional (INT—Pascale)	Melanoma	578 (total) 80% (training set) 20% (test set) 117 (ex. validation)	Ipilimumab, nivolumab, pembrolizumab	More clinical variables: CLICAL signature	Cox model	ML; survival analysis RSF Survival analysis (statistics) stratification in five risk groups	Time-dependent AUC-ROC at 36 month 0.81 Comparison between five risk groups: $P = 0.001$
Liu et al., 2021[125]	Retrospective	Post-marketing surveillance dataset	Gynaecological cancer	117 (total) 70%	Camrelizumab, sintilimab, toripalimab	15 clinica features	RFE	ML Classification; ensemble method kNN-stacker	Death (simple-XGB) Acc. 0.9706; Rec. 1; Prec. 0.75; F1 0.8571

Continued

Table 5. Continued

Authors and years	Type of study	Data source	Type of cancer	No. of patients	Therapy	Biomarkers	Feature selection	Developed model	Outcomes results
				(training) 30% (test)				XGB CB Ensemble method GAIN-blender	General AE (GAIN-Blender) Acc. 0.7353; Rec. 0.7273; Prec. 0.5714, F1 0.64 Organ AE (kNN-CB) Acc. 0.6176; Rec. 0.7222, Prec. 0.619 F1 0.6667 Binary RECIST (KNN-Stacker) Acc. 0.7059; Prec. 0.8947; Rec. 0.68; F1 0.7727
Wu et al., 2022[126]	Prospective	Institutional (TAHSYUG)	HCC	35 (total) 70% (train) 30% (test)	Anti-PD-1	18 key microbial biomarkers; *Phascolarcto bacterium* *Odoribacter* *Megasphaera* *Aggregatibacter* *Escherichia* *Shigella* *Gemmiger* *Cloacibacillus* *Ruminococcus* *Klebsiella* *Haemophilus* *Coprococcus* *Parabacteroides* *Faecalibacterium* *Barnesiella* *Parasutterella* *Dorea* *Lactobacillus*	Boruta algorithm	ML Classification: RF	Responders versus non-responders AUC 0.837
Harel et al., 2022[127]	Retrospective	Biobank (Indivumed, Germany), Institutional (ISMC)	NSCLC	143 (total) 72 + 36 (training) 35 (test)	Immunotherapy (anti-PD-1 or anti-PD-L1 alone or in combination with chemotherapy)	Gender, age, CXCL10, CXCL8 (from an ELISA -based array including 840 proteins)	Limit of detection according to RayBiotech definitions. Features with FDR <0.15 were selected to generate the predictor. Clustering analysis was carried out following Z-score normalisation	ML: classification XGB algorithm for prediction of response to immunotherapy	ORR AUC = 0.79; sen. = 0.78, spec. = 0.65, PPV = 0.70 and NPV = 0.73

Acc, accuracy; AE, adverse event; AFP, alpha-fetoprotein; AP-HM, Assistance Publique-Hopitaux de Marseille; AUC-ROC, area under the receiver operating characteristic curve; BNM, Bayesian network mode; BOR, best overall survival; CLICAL, clinical categorisation algorithm; CRCM, Cancer Research Center of Marseille; CRP, C-reactive protein; CTLA-4, cytotoxic T-lymphocyte antigen 4; CXCL10, C-X-C motif chemokine ligand 10; CXCL8, C-X-C motif chemokine ligand 8; CXCL9, C-X-C motif chemokine ligand 9; DCE, dynamic contrast-enhanced; DCR, disease control rates; DL, deep learning; dNLR, derived neutrophil-to-lymphocyte ratio; EESC, esophageal squamous cell carcinoma; Ex. Val, external validation; GAC, gastric adenocarcinoma; GBM, gradient boosting machine; GEO, Gene Expression Omnibus; GLM, generalised linear models; GRI, Gustave Roussy Institute; GSEA, gene set enrichment analysis; Hb, haemoglobin; HCC, hepatocellular carcinoma; HNSCC, head and neck squamous cell carcinoma; IO, immunotherapy; k-NN, k-nearest neighbours; LASSO, least absolute shrinkage and selection operator; LDH, lactate dehydrogenase; LR, logistic regression; MALDI, deep matrix-assisted laser desorption/ionization; ML, machine learning; MSKCC, Memorial Sloan—Kettering Cancer Center; nlr, negative likelihood ratio; NN, neural network; NPV, negative predictive value; NSCLC, non-small-cell lung cancer; OS, overall survival; PCA, principal component analysis; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; PFS, progression-free survival; PLR, positive likelihood ratio; PPV, positive predictive value; PS, performance status; PSEA, protein set enrichment analysis; RCC, renal cell carcinoma; RF, random forest; RT, regression tree; SL, SuperLearner; SVM, support vector machine; TCGA, The Cancer Genome Atlas; TCR, T-cell receptor; TRAE, treatment-related adverse events; WBC, white blood cell; XGB, eXtreme Gradient Boosting.

[a]Diagnostic Cortex™ - a platform that employs modern machine learning techniques to design deep learning clinical trials for lung cancer (https://www.biodesix.com/newsroom/press-releases/biodesixs-diagnostic-cortex-platform-used-three-studies-presented-sitc)

[b]JADBio (Just Add Data Bio)- a fully automated machine learning (AutoML) system (https://jadbio.com/)

https://doi.org/10.1016/j.annonc.2023.10.125

appeared to be recurrent to most of the algorithms e.g.: texture, shape, volume, kurtosis, heterogeneity, brightness, dynamics, gray-level co-occurrence matrix and gray-level size zone matrix (Figure 2).

Pathomics

Table 4 summarises eight retrospective studies that use pathomics data, 50% ($n = 4$) of which were published in 2022. In the Supplementary Table S3, available at https://doi.org/10.1016/j.annonc.2023.10.125, we report seven further pathomic studies that use AI to predict molecular features with therapeutic impact. Eighty-seven percent of the studies have been conducted on NSCLC ($n = 3$), melanoma ($n = 2$), or including both cohorts with other cancer types cohort ($n = 2$). Seventy-five percent ($n = 6$) used DL methods, mainly CNN. Codes are available for 38% ($n = 3$) of studies. In pathomics, standard ML methods were used to (i) carry out segmentation and classification, (ii) create a prognostic score, or (iii) validate a developed signature. As an example, Johannet et al.[102] utilised a two-stage Inception-V3 CNN for segmentation and multivariable classification, integrating CNN predictions and clinical data to predict objective response rate in melanoma patients (AUC = 0.8) and categorise patients into high versus low risk of progression ($P = 0.003$). As for the third approach, Chen et al.[104] used an image analysis system (inForm) to create a multidimensional tumour-infiltrating immune cells (TIICs) signature for gastric cancer. The developed signature was used in different ML models to identify responders to ICI, and the best-performing model (AdaBoost) achieved an AUC of 0.85 revealing the inadequacy of a single biomarker.

Table 4 and Figure 3 summarise the most frequent biomarkers constructed from histological images. Among these, PD-L1 and TILs were the most explored with a DL approach (Table 4). In particular, Harder et al.[98] identified image-based signatures in malignant melanoma that may serve as diagnostic tools for ipilimumab, using a decision tree model based on CD8+ tumour-infiltrating lymphocytes. Park et al.[99] revealed three immune phenotypes (inflamed, immune-excluded, and immune-desert) as potential biomarkers for predicting response to ICI in advanced NSCLC. Related to PD-L1, Hu et al.[103] demonstrated the potential of CNN in predicting anti-PD-1 response in melanoma and lung cancer patients. Choi et al.[105] showcased the efficacy of an AI-powered tumour proportion score (TPS) analyser to decrease interobserver variation among pathologists and increase accuracy in therapeutic response prediction for NSCLC. Finally, in a subset of studies, digital pathology datasets with matched molecular data have permitted exploratory work in predicting molecular biomarkers (e.g. driver gene mutation[106,107]) from haematoxylin and eosin (H&E) images.

Real-world data

This group includes standard ML methods applied to the analysis of RWD, found in 20 studies (see Table 5).

Thirty percent ($n = 6$) of included studies are conducted on NSCLC patients, followed by melanoma (20%, $n = 4$) and multicancer (15%, $n = 3$) cohorts. Most of the studies (16 out of 20, 80%) from Table 5 used supervised classification (75%, $n = 12$ standard ML; 12.5%, $n = 2$ DL; and 12.5%, $n = 2$ both standard ML and DL) to predict responders versus non-responders. Two (10%) studies used ML survival analysis exploiting continuous outcomes (OS and PFS). The main goal of implementing ML methods was to valorise these data by capturing nonlinear interactions between used features, to achieve an improved prediction of ICI efficacy by using easily accessible and cost-efficient data. For example, Gupta et al.[121] developed a Bayesian network model using sociodemographic characteristics, tumour characteristics, and prior treatment types to predict OS in RCC patients who received nivolumab therapy. Their model outperformed the International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) risk score,[128] a widespread prognostic model for targeted therapy in RCC in the external validation cohort, achieving a mean AUC for OS at 12 months of 0.76, compared to IMDC's AUC of 0.69. For example,[112] one study used radiology text reports to estimate RECIST-defined outcomes and PFS using a fully connected DL. The developed model, using only the text reporting the data as input, correctly predicted RECIST PFS data in 82% of cases within 2 months. They further compared the model's prediction to the manual review of trained medical oncologists and achieved similar results. Biomarkers identified for RWD AI-based algorithms (Figure 2) can be divided into three groups: (i) patient, (ii) treatment, and (iii) cancer-related. Among patient-related RW biomarkers: NLR, LDH, haemoglobin, C-reactive protein, performance status (PS), platelets, body mass index, patient reported outcome, antibiotics, steroids, age, sex, and gut microbiome were identified. Treatment information [e.g. ICI class type, line and type of IO-based therapy (combined or not with CT)] have been used to feed the algorithm. Finally, the most frequent cancer-related biomarkers were tumour location, histotype, stage, tumour burden, liver, and central nervous system metastasis.

Multimodal data

Lastly, we identified five retrospective studies that integrated multimodal data (minimum three modalities) to predict ICI efficacy, Table 6. They all used a two-step algorithm development: firstly, they created a signature based on different data combinations and, subsequently, they provided a method to predict the patients' survival.

Four out of five (80%, $n = 4$) used DL to integrate the data. The NSCLC study[129] demonstrates the value of multimodal integration in predicting ICI efficacy. The model with multimodal data, integrated using the intermediate fusion,[130] outperformed all unimodal models including TMB, radiomics, and pathomics on the same dataset, achieving an AUC of 0.80 even when data were unavailable for some patients. Similarly, the integrated DL model proposed by Yang et al.[131] which incorporates serial radiomics,

laboratory data, and clinical information of patients with advanced NSCLC, demonstrates an impressively high prediction capability, surpassing traditional RECIST evaluations. Moreover, AI-driven discovery tools have further helped identify novel immune phenotypes, such as the LAG-3+CD8+ T-cell population described by Shen et al.[132] from immune checkpoint blockade-treated patients. This distinct phenotype serves as a prognostic marker for poorer outcomes in patients with melanoma and urothelial carcinoma.

Sophisticated AI models leveraging on medical imaging data were developed by Park et al.,[133] which used quantitative flow cytometry and RNA-seq immune profiles derived from [^{18}F]2-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for predicting CytAct scores in lung adenocarcinoma effectively. Similarly, Mu et al.[134] combined genomics with PET radiomics and clinical data in different cohorts, validating the algorithm in an ICI cohort of 149 patients. However, all the selected studies featured small-volume datasets (<900 patients) despite high-dimensionality, and critically only one validated results using an external dataset. The code is available for two studies (40%).

DISCUSSION

This review critically appraises state-of-the-art AI approaches (ML and DL) that strive to discover data-driven biomarkers predicting ICI treatment efficacy in pan-cancer settings. To the best of our knowledge, this is the first systematic review of AI methodologies applied to immuno-oncology that includes different data modalities across all cancer types. We appraised 90 identified studies across four major data modalities in cancer: genomics (including transcriptomics and epigenomics), histopathology (pathomics), radiology (radiomics), real-world and multimodal data describing the datasets, methodologies, biomarkers, and results. We report that new AI methodology use for ICI efficacy prediction is rising, with 80% ($n = 72$) of included articles published between 2021 and 2022, most (85% $n = 84$) being retrospective. Biomarker discovery traditionally involves analysis of potentially informative qualitative (such as tissue morphology) or quantitative features (such as genomics, blood exams) and their association with clinical outcomes.[135] AI/ML/DL methodologies permit exploration of high-throughput data, with nonlinear relationships among variables to better select features that independently impact prognosis (e.g. LASSO analysis); this could lead to the creation of models that can reconsider previously discarded biomarkers. In this review we identified the potential of these 'new' technologies in all data categories.

Identifying genes associated with ICI response is challenging because of the increasingly high-dimensionality of multi-omics datasets owing to improvements in sequencing technologies. ML offers solutions to handle high-dimensional data, including supervised ML-based gene selection, unsupervised clustering, and DL models. Some studies developed signatures using non-ICI therapies and validated them with ICI-treated patient cohorts, implying a prognostic rather than predictive nature. Moreover, for the development and validation of AI, structured public data storage is needed. Existing public platforms such as The Cancer Imaging Archive (TCIA),[136] TCGA,[71] and GEO[72] are good examples of cancer open data storages, where researchers can access large pools of data needed for research, while Cbioportal[73,74]offers an open-source resource for the interactive exploration of genomics datasets. Actually, many papers on genomics (Table 2) were based on these public platforms. Despite their strengths, these platforms usually contain only small patient cohorts with multiple data modalities and without complete outcome information and hence high-dimensional censored data. Finally, geometric network analysis can be used to utilise gene pathway information instead of the aforementioned apporaches.[137]

Radiomics is one of the most represented fields in prediction of cancer efficacy to ICI enabled by the use of radiological images to feed emerging AI technologies. Two systematic reviews/meta-analyses[138,139] have been conducted on radiomic biomarkers: one focused on NSCLC patients treated with anti-PD-(L)-1[138] and the other included all cancer types but excluded studies using DL methods.[139]

Advancements in digitalisation of histological glass slides into whole slide images (WSI) have enabled pathomics to rapidly develop. Computational analysis of digitalised histology slides can extract valuable information for improving clinical decision making in cancer immunotherapy (i.e. histology, PD-L1 TPS images). However, immune-pathology is relatively unexplored, with most studies focusing on diagnosis or identification of biomarkers that could be associated with benefit to ICI (e.g. MSI,[140] PD-L1 status,[141] and inflammatory genes[142]). DL shows great potential to extract biomarkers directly from slides and allows for predictive biomarker discovery; however, validation and implementation of these models in clinical practice is challenging due to the rarity of digitalised datasets. Digital capacity building and introducing them into routine histopathological workflows is therefore necessary.

In the real-world environment, ML models can be used to process vast amounts of information that usually remain unprocessed,[112] e.g. EHR. Compared to other data modalities, RWD is easy to access, routinely collected, and is easier to handle because it does not require complex preparations such as specific annotation from experts or image pre-processing. In addition, neural language processing or text mining can be used to extract more information from EHRs, but data accuracy control is necessary to fully trust the results from automated processes, such as the use of ChatGPT. RWD curation is crucial for algorithm validation and extensive use as it has a higher risk of being originally inaccurate compared to other data types.

AI biomarkers from RWD showed consistent results with conventional statistical methods, confirming the significance of factors such as Eastern Cooperative Oncology Group PS, NLR, and LDH (Table 5). This strengthens the trust

Table 6. Studies using multimodal features and artificial intelligence to predict immunotherapy response in cancers

Multimodal data									
Park et al., 2020[133]	Retrospective	TCIA, GEO, Institutional (SNUH), TCGA	Lung adenocarcinoma	211 (total) 93 (training), 59 (ex. validation1), 59 (ex. validation 2 IO set)	Nivolumab, pembrolizumab, atezolizumab	CytAct	DNN	DL; creation of a score—CytAct Survival analysis (statistics); univariant LR and K-M; stratification of patient using score	Partial response (Y/N): AUC 0.88 PFS: HR = 0.25, P = 0.001 OS: HR = 0.18, P = 0.004
Mu et al., 2020[134]	Retrospective	Institutional (SPH, HBMU, HMU, HLM)	NSCLC	897 (total) 429 (training) 187 (test) 65 (ex. validation) 149 (IO validation set) 67 (TKI validation set)	Anti-PD-L1, anti-PD-1	EGFR mutation status score	DNN	DL: creation of EGFR-DLS mutation status score Survival analysis: (statistics): K-M; stratification of patient using EGFR-DLS score	PFS: HR = 2.33, P < 0.001
Vanguri et al., 2022[129]	Retrospective	Institutional (MSK)	NSCLC	247 (total; 10-fold cross-validation) 52 (pathology validation) 46 (radiology validation)	Anti-PD-(L)1 +/- CTLA-4 combination	DyAM risk, dNLR, albumin, brain metastases, combined treatment	Based on expertise, L-regularisation	ML and DL multimodal dynamic attention with masking DyAM model Survival analysis: Cox proportional hazard (using DyAM score as an input) K-M: stratification between low- and high-risk score	BOR: 10-fold cross-validation AUC = 0.80 (95% CI 0.74-0.86) PFS: C-index 0.74
Shen et al., 2022[132]	Retrospective	Clinical trials (NCT0102423, NCT01295827, NCT01621490, NCT01844505, NCT01927419, NCT01928394, NCT02083484, NCT02553642, NCT03122522)	Melanoma, urothelial carcinoma	282 (total) 188 (training) 94 (test)	Anti-PD-1, anti-CTLA-4	LAG-3 expression, PD-L1, TMB		ML: unsupervised clustering: survClust for LAG 3 signature creation Survival analysis (statistics), stratification	Test cohort OS: P < 0.001 Test subset (Imonotherapy) P < 0.001 PFS: P = 0.004
Yang et al., 2021[131]	Retrospective	Clinical trials (CheckMate-870, CheckMate-078, OAK), institutional (SLCC)	NSCLC	200 (total) (fivefold cross-validation)	Anti-PD-1 /PD-L1	Multidimensional serial information: Radiomics, clinical, laboratory data	Clinical	DL Classification; SimTA Survival analysis (statistics): stratification in low- and high-risk groups using model pred.	Responders and non-responders AUC 60 days 0.77 AUC 90 days 0.80 Low- versus high-risk groups: Median PFS: P < 0.01 Median OS: P < 0.01

Acc, accuracy; AdaBoost, Adaptive Boosting; AE, adverse event; AFP, alpha-fetoprotein; AP-HM, Assistance Publique-Hopitaux de Marseille; ASR, all-subsets regression; ATHZU, Affiliated Tumor Hospital of Zhengzhou University; AUC-ROC, area under the receiver operating characteristic curve; BCC, basal cell carcinoma; BNM, Bayesian network model; BOR, best overall survival; CB, CatBoost; CCL5, (C-C Motif chemokine ligand 5); CD4, cluster of differentiation 4; CD8, cluster of differentiation 8; CDARS, clinical data analysis and reporting system; CRCM, Cancer Research Center of Marseille; CRP, C-reactive protein; CTLA-4, cytotoxic T-lymphocyte antigen 4; CXCL10, C-X-C motif chemokine ligand 10; CXCL8, C-X-C motif chemokine ligand 8; CXCL9, C-X-C motif chemokine ligand 9; CytAct, cytolytic activity score; DCB, durable clinical outcome; DCE, dynamic contrast-enhanced; DCR, disease control rates; DL, deep learning; DLS, deep learning score; dNLR, derived neutrophil-to-lymphocyte ratio; DyAM, deep attention-based multiple-instance learning model with masking; EGFR, epidermal growth factor receptor; ELISA, enzyme-linked immunosorbent assay; ESCC, esophageal squamous cell carcinoma; Ex. Val, external validation; FDR, false discovery rate; FRRS, ferroptosis-related risk score; GAC, gastric adenocarcinoma; GAIN, generative adversarial imputation nets; GBM, gradient boosting machine; GEO, Gene Expression Omnibus; GLM, generalised linear models; GRI, Gustave Roussy Institute; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; Hb, haemoglobin; HBMU, Fourth Hospital of Hebei Medical University; HCC, hepatocellular carcinoma; HLM, Lee Moffitt Cancer Center and Research Institute; HMU, Fourth Hospital of Harbin Medical University; HNSCC, head and neck squamous cell carcinoma; HR, hazard ratio; ICI, immune checkpoint inhibitor; IMDC, International Metastatic Renal Cell Carcinoma Database Consortium; INT, Istituto Nazionale Tumori; IO, immunotherapy; ISMC, Israel Sheba Medical Center; K-M, Kaplan—Meier; k-NN, k-nearest neighbours; LAG-3, lymphocyte-activation gene 3; LASSO, least absolute shrinkage and selection operator; LDH, lactate dehydrogenase; LR, logistic regression ; MALDI, deep matrix-assisted laser desorption/ionisation; MCC, Matthew's correlation coefficient; MGH, Massachusetts General Hospital; MHC, major histocompatibility complex; ML, machine learning; MLP, multilayer perceptron; MSCC, Mass General Cancer Center; MSKCC, Memorial Sloan—Kettering Cancer Center; NDB, no durable benefit; NDB-LS, long-term survival with no durable clinical benefit; NKI, Netherlands Cancer Institute; nlr, negative likelihood ratio; NLR, neutrophil-to-lymphocyte ratio; NN, neural network; NPV, negative predictive value; NSCLC, non-small-cell lung cancer; ORR, objective response rate; OS, overall survival; PBIMs, peripheral blood mononuclear cells; PCA, principal component analysis; PD-1, programmed cell death protein 1; PDAC, pancreatic ductal adenocarcinoma; PD-L1, programmed death-ligand 1; PFS, progression-free survival; PLR, positive likelihood ratio; PPS, poor prognosis signature; PPV, positive predictive value; Pre, precision; PS, performance status; PSEA, protein set enrichment analysis; RCC, renal cell carcinoma; Rec, recall; RECIST, Response Evaluation Criteria in Solid Tumors; RF, random forest; RFB, random forest and Boruta; RFE, recursive and feature elimination; RF-SRC, fast unified random forests for survival, regression, and classification; RSF, random survival forest; RW, real world; Sen, sensitivity; SimTA, simple temporal attention; SLCC, Shanghai Lung Cancer Center; SNUH, Seoul National University Hospital; Spec, specificity; SPH, Shanghai Pulmonary Hospital; SVM, support vector machine; SVM-RFE, support vector machine-recursive and feature elimination; TAHSYUG, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou; TCGA, The Cancer Genome Atlas; TCIA, The Cancer Imaging Archive; TCR, T-cell receptor; TKI, tyrosine kinase inhibitor; TMB, tumour mutational burden; TRAE, treatment-related adverse events; UGHH, University General Hospital of Heraklion; Umap, Uniform Manifold Approximation and Projection; VUMc, Vrije University Medical Center; EMC, Erasmus University Medical Center; WBC, white blood cell; XGBoost, eXtreme Gradient Boosting; Y/N, yes/no; YCC, Yale Cancer Center; YSPORE-SC, Yale SPORE in Skin Cancer.

in AI algorithms when appropriately developed, validated, and explained.[143] Integrating multimodal data remains the 'cherry on top' and AI has been recognised for its ability to extract and combine complementary contextual information across different modalities to improve decision making in oncology.[13,130,135,144]

Recently, three different multimodal data integrations have been reported: (i) early fusion—the creation of a joint representation from raw data or features at the input level before feeding it to the model, (ii) late fusion—training a separate model for each modality and aggregating the predictions from individual models at the decision level, and (iii) intermediate fusion—unimodal features are initially processed separately before a fusion step and subsequent analysis of the fused representation.[130] Compared to traditional statistical methods and unimodal models, integrating multimodal data enables the creation of meta-biomarkers by merging known single/original biomarkers and discovering new ones from diverse sources. Multimodal models have shown potential to improve performance, with an increase in AUC from 0.65 to 0.80 when compared to unimodal models on the same dataset.[129] Data integration also offers several advantages, the opportunity to develop personalised medicine by considering the phenotype, the genotype and exposome (e.g. behavioural data with new types of monitoring using wearable devices); future studies will lead to a more in-depth AI phenotyping.

To ensure that multimodal models can ultimately be translated to clinic, it is essential that such models are trained on data available in the clinical setting—clinical features, imaging and simple molecular tests. It is important that the field does not generate multimodal models that are reliant on expensive and technically difficult data sources (e.g. whole-genome sequencing) that are unlikely to be available to make predictions in the clinic. Building multimodal models reliant on simple data types increases the possibilities for clinical deployment in the clinical setting, permitting wide-scale delivery of precision medicine.

Despite its systematic design, our study has limitations such as excluding methodologically important studies not directly related to ICI treatment, small sample size publications (<100), and studies that only used AI for feature selection. To address this, Supplementary Table S3, available at https://doi.org/10.1016/j.annonc.2023.10.125, was added to include some of the excluded publications meeting these criteria.

In spite of its promise and potential, there are ongoing challenges to be solved before AI can be integrated into biomarker discovery and implementation in immuno-oncology clinical practice:

Data scarcity and structure: One major limitation to develop ML methodologies is the requirement of large amounts of high-quality data. Oncological datasets are inherently complex, often high-dimensional, incomplete, biased, heterogeneous, and noisy.[144] To achieve good algorithmic performance, large volumes of high-quality data are essential necessitating time-consuming curation and heavy pre-processing.

Models' heterogeneity: Currently, numerous different ML-based models are available causing difficulties for researchers to decide which methodology to use. When choosing the model, researchers should consider the available data (in volume and diversity), and be aware that in case of low-volume structured data classical ML methodologies are sufficient.[145]

Standardisation, protocols, and guidelines in conducting the research and results reporting: Medical AI systems must undergo evaluation via randomised controlled trials, the current gold standard.[146] Reporting guidelines like CONSORT-AI[147] and SPIRIT-AI[148] now include AI-specific recommendations, and the ML-CLAIM[149] checklist aims to improve transparent reporting. However, still many studies lack clear documentation on training, test and validation cohort selection, methodological development, validation, comparison with standard methods, and its plan for clinical integration. Initiatives like ESMO-GROW may offer updated guidance for RW evidence studies in oncology, including novel ML methods. As the field advances, and more ML methods are tested in clinical trials, following these guidelines is mandatory.

Biomarkers undergo approval procedures from regulatory agencies, an important step but in need of standardisation. Similarly, to integrate AI models and the tools built upon them as medical software devices, it is necessary to develop a generalised process for their certification. This is essential for responsible deployment of AI in the clinic and is a requirement to meet the standards of good medical practice.

Models' interpretability and prediction explainability: Finally, a major hurdle for ML adoption in medicine is the lack of trust in the models.[145,146,150] Algorithmic complexity can be illuminated with XAI methods, which show results in an intuitive way making them easy for clinicians and patients without AI experience to understand. XAI also allows us to interrogate model design and predictive decision-making strategies to enhance biological discovery, e.g. revealing molecular mechanisms underlying primary resistance to immunotherapy. Most of the current studies, as well as studies included in this review, do not include an interpretation or explanation of the models and their prediction in their work. This critically important step, XAI, should be necessary in any ML pipeline. XAI in DL models, e.g biomarker prediction from images, is more challenging but ongoing methodological innovations, for example with saliency mapping techniques,[32] offer a solution and must be supported if such tools are to be clinically implemented.

Models' generalisability and robustness: ML models for clinical practice need exposure to diverse data sources from multiple medical centres to prevent algorithm bias caused by differences in data structure, staining intensity, patient demographics, sex, etc. External validation of models is being increasingly reported in studies (50% of the studies included in this review), mainly for the studies that are using data from publicly available data sources, i.e. genomics studies.

Study design: The studies reviewed lacked clearly predefined endpoints and pipelines for immediate

incorporation of predictive algorithms into clinical practice. To advance, future studies should be designed as prospective clinical trials with accurate pre-planned AI methodology, such as the I3LUNG trial (NCT05537922). More data-driven observational studies are needed, particularly for biomarker-based discovery, like the APOLLO 11 trial (NCT0555096).

CONCLUSIONS

In this systematic review, we confirmed an increasing use of AI in order to discover predictive biomarkers for the efficacy of ICIs in various cancers, an approach that can be expanded to other fields such as the efficiency of chemotherapy or targeted therapies. AI methodologies have provided new insights from complex data, but developing AI-based 'software biomarkers' is hindered by retrospective datasets, diverse AI methodologies, and opaque decision making. While these studies provide some hypothesis-generating insights, direct clinical implementation is limited. To create an interpretable and responsible AI tool, large-scale prospective validation studies are necessary. Such tools are critical for IO, where new meta-biomarkers are needed to predict response.

ACKNOWLEDGEMENTS

The panel would like to acknowledge the work of Klizia Marinoni and Fiona Perdomo, from the Scientific and Medical Division at ESMO, for the editorial assistance, and Giovanni Scoazec's contribution for the English language review.

FUNDING

None declared.

DISCLOSURE

AP declares training of personnel role for AstraZeneca and Italfarma; advisory board role for BMS, travel grant from Janssen; and speaker's engagement for Roche. FT declares full or part-time employment with Politecnico di Milano as researcher and lecturer for high-level education; and as cofounder of MLcube. CG declares advisory board role for Amgen, Roche, Sanofi, Takeda; speaker's engagement from AstraZeneca, Bristol Myers Squibb, Eli Lilly, Merck-Sharp-Dohme, and Novartis; institutional funding from AstraZeneca and Bristol Myers Squibb; institutional research grant from Italian Ministry of Health; clinical trials PI for AstraZeneca and Roche; and non-remunerated activities as member or AIOM, AIOT, ASCO, FONICAP, IASLC, and ISLB. SER declares speaker's engagement from Amgen, BMS, Astellas, and GSK; writer's engagement from BMS and Amgen; and travel and accommodation from Janssen and MSD. LC-B declares full time employment with ESMO as Scientific and Medical division Fellow. JD declares consultant role for MJH Life Sciences at ASCO Clinical Congress Consultants at ASCO 2022. ATP declares advisory board role for Abbvie, Elevar, Prelude Therapeutics, and Privo; institutional research grant from Abbvie and Kura; and non-renumerated activity as PI for Adenoid Cystic Carcinoma Research Foundation and Cancer Research Foundation. GLR declares advisory board role for AstraZeneca, BMS, MSD, Novartis, Pfizer, Roche, and Sanofi; speaker's engagement from Italfarnaco; institutional funding as local PI for Amgen, AstraZeneca, BMS, Celgene, GSK, MSD, Novartis, Roche and Sanofi. CP declares personal fees from Italfarmaco, AstraZeneca, BMS, Merck Sharp and Dohme, and Janssen; and institutional funding from Novartis. ST declares speaker's engagement from IDEA Pharma, Merck, MSD, Roche, and Ventana; institutional funding from Andy Quick Charitable fund, Complete genomics, CRUK, CRUK training and career development board, CRUK Welcome Trust, Harry J Lloyd Charitable Trust Career Development Award, NIHR, RMH/ICR/BRC/Imperial AHSC/Faculty of Medicine, Rosetrees Trust, The Francis Crick Institute, The Robert McAlpine Foundation, and Ventana. MK declares advisory board role for Bayer, MSD, Pierre Fabre, Servier; speaker's engagement for BMS, Merck; institutional funding from Amgen, BMS, Merck, Nordic Farma, Novartis, Pierre Fabre, Servier; institutional research grant from Bayer, Bristol Myers Squibb, Personal Genomics Diagnostics, Pierre Fabre, Roche, Servier, Sirtex; trial chair for Servier; non-remunerated activities for Leadership Role for Dutch Colorectal Cancer Group and ESMO; advisory role for KWF, Patient representative organization (Kanker.nl), ZiNNL; and non-remunerated activity as ESMO faculty member. SD declares advisory board role for AstraZeneca, Besins Healthcare, Cellectis, Elsan, Isis/Servier, Novartis, Orion, Pierre Fabre, Rappta, Sanofi; speaker's engagement for AstraZeneca, Exact Sciences, Gilead, Lilly, MSD, Pfizer, Seagen; institutional funding as steering committee member for AstraZeneca, BMS, Pfizer, Puma, Sanofi, Roche Genentech; institutional funding as local PI for Orion; institutional funding as coordinating PI for Taiho; institutional funding from GE; non-remunerated activity as PI for European Commission; and non-remunerated activity as Board of Directors member for SFSPM. JNK declares advisory board role for DoMore Diagnostics, Owkin, and Panakeia; and speaker's engagement from Eisai, Fresenius, and MSD. FGMDeB declares speaker's engagement from Ambrosetti, Amgen, BMS, Dephaforum, ESO, Healthcare Research & Pharmacoepidemiology, Incyte, Merck Group, MSD, Nadirex, Pfizer, Roche, Sanofi, Seagen, and Servier; consultation/advisory role from Mattioli 1885, AstraZeneca, BMS, EMD Serono, Incyte, Menarini, MSD, NMS Nerviano Medical Science, Novartis, Pierre Fabre, Roche, Sanofi, and Taiho; funding as coordinating PI from Basilea Pharmaceutica International AG, BMS, DAICHI SANKIO Dev. Limited, Exelixis Inc, F. Hoffman-LaRoche Ltd, Ignyta Operating INC, IQVIA, Janssen-Cilag International NV, Kymab, LOXO Oncology Incorporated, MedImmune LCC, Merck KGaA, Merck Sharp & Dohme Spa, MSD, Novartis, Pfizer, and Tesaro; and think tank funding from MCCann Health. MCG declares speaker's engagement from AstraZeneca, ecancer, GrupoPacifico-Secretaria Tecnica ICAPEM/AstraZeneca, Medscape, MSD, MSD Italia, S.O.S S.r.l, WebMD, WebMD Oncology/Takeda; consultation/advisory role from AstraZeneca/MedImmune,

Daiichi Sankyo, AstraZeneca Poland, AstraZeneca UK, Bayer Healthcare Pharmaceuticals, Blueprint Medicines, Daiichi Sankyo, Eli Lilly, GlaxoSmithKline, Incyte, Mirati Therapeutics, MSD, Pfizer, Regeneron Pharmaceuticals, Roche, Sanofi/Prex, Sanofi Genzyme corporation, Seagen International GmbH, and Takeda; institutional funding as local PI for Phase III studies from Amgen, AstraZeneca, Bluprint, BMS, GlaxoSmithkline, Incyte Corporation, IPSEN Bioscience, MedImmune LCC, Novartis, Roche, and Sanofi; institutional funding as local PI for Phase II studies from AstraZeneca, Celgene Corporation, Daiichi Sankyo Development, MedImmune LCC, Merck KGaA, Merck Serono, Spectrum Pharmaceuticals, and Turning Point Therapeutics; institutional funding as local PI for Phase I studies from Exelixis Inc., Janssen; institutional funding as local PI from Bayer, Janssen, Merck, Pfizer, Otsuka Pharmaceutical Italy; institutional funding as Coordinating PI from AstraZeneca, MSD; steering committee member of the MK-3475 KN671 Steering Committee (KEYNOTE-671) from MSD, MK-3475 KN671 Steering Committee from MSD, Steering Committee ML41118 from Roche, and steering committee of Pfizer Global Lung Cancer Educational Programme; personal financial interests as PACIFIC-R Scientific Committee from AstraZeneca, PACIFIC-R Global Scientific Committee from AstraZeneca, Pacific 6 International Coordinating Investigator from AstraZeneca, steering Committee member and Co-chair at the AstraZeneca Lung Cancer Summit 2019, GSK-Garassino-ZEAL Steering Committee 2020-23, GSK Lung Cancer Global Council, Jannesen Scientific Advisory Board and Therapeutic Area Steering Committee Meeting on Lung Cancer, Seattle Genetics Lung Cancer Platform Study; travel, accommodations and expenses from AstraZeneca, Pfizer, and Roche; non-remunerated activities as principal investigator of AO Spedali Civili Brescia, Eli Lilly Studio TYME trial, European Thoracic Oncology Platform (ETOP) phase III trial, GOIRC Studio CHANCE trial, GOIRC RAME trial, GUSTAVE-ROUSSY PARIGI LIPI TRIAL, Istituto dei Tumori Pascale MILES 5 trial, Istituto dei Tumori Pascale IND227 trial, Istituto dei Tumori Pascale Beverly trial, Istituto Nazionale dei Tumori Progetto Timoma trial, Istituto Nazionale dei Tumori APOLLO trial, Istituto Nazionale dei Tumori Bando finalizzata Mesotelioma trial, Istituto Nazionale dei Tumori POST-ALK trial, Istituto Nazionale dei Tumori TERAVOLT trial, Istituto Nazionale dei Tumori FAME trial, MSD People, Pfizer STYLE trial, and Sant'Orsola Malpighi Creta trial; non-remunerated activities as member of AIOM, AIOT, ASCO, EMA Scientific Advisory Group, ESMO, IPOP, TUTOR, WCLC, Women for Oncology Italy; and non-remunerated activities as Scientific Programme Committee of AACR. GP declares full time employment from ESMO. ALGP declares personal financial interests for stocks/shares as cofounder of Agade s.r.l. and AllyArm s.r.l; institutional financial interests for co-funding of the Professorship Chair by Associazione Univerlecco and Congregazione Suore Infermiere dell'Addolorata; institutional financial interests for licensing/royalties as co-inventor of Patent No. 10201900000848 (Load compensation device, registered with Politecnico di Milano, surrendered in January 2023 with a royalty return plan), Patent No. 102018000006950 (System for detection and kinematic monitoring of body movements in water, and relative method, registered with Politecnico di Milano and Fondazione per la Ricerca Scientifica Termale; PCT/EP2019/067618) and Patent No. 102017000027918 (Device for controlled assistance of the grip; PCT/IB2018/051652); and non-financial interests as a member of the International Conference of Rehabilitation Robotics, member of the Institute of Electrical and Electronics Engineers (IEEE), IEEE Robotics and Automation Society, IEEE Engineering in Medicine and Biology Society and IEEE Women in Engineering. All other authors have declared no conflicts of interest.

REFERENCES

1. Hodi FS, Chiarion-Sileni V, Gonzalez R, et al. Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. *Lancet Oncol*. 2018;19(11):1480-1492.
2. Burtness B, Harrington KJ, Greil R, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet*. 2019;394(10212):1915-1928.
3. Fradet Y, Bellmunt J, Vaughn DJ, et al. Randomized phase III KEYNOTE-045 trial of pembrolizumab versus paclitaxel, docetaxel, or vinflunine in recurrent advanced urothelial cancer: results of >2 years of follow-up. *Ann Oncol*. 2019;30(6):970-976.
4. Bellmunt J, de Wit R, Vaughn DJ, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N Engl J Med*. 2017;376(11):1015-1026.
5. Escudier B, Sharma P, McDermott DF, et al. CheckMate 025 randomized phase 3 study: outcomes by key baseline factors and prior therapy for nivolumab versus everolimus in advanced renal cell carcinoma [Figure presented]. *Eur Urol*. 2017;72(6):962-971.
6. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Five-year outcomes with pembrolizumab versus chemotherapy for metastatic non—small-cell lung cancer with PD-L1 tumor proportion score ≥ 50. *J Clin Oncol*. 2021;39(21):2339-2349.
7. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non—small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135.
8. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1—positive non—small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823-1833.
9. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced non-squamous non-small cell lung cancer. *N Engl J Med*. 2015;373(17):1627-1639.
10. Rodríguez-Abreu D, Powell SF, Hochmair MJ, et al. Pemetrexed plus platinum with or without pembrolizumab in patients with previously untreated metastatic nonsquamous NSCLC: protocol-specified final analysis from KEYNOTE-189. *Ann Oncol*. 2021;32(7):881-895.
11. Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US food and drug administration (FDA) approvals of immune checkpoint inhibitors. *J Immunother Cancer*. 2019;7(1):278.
12. Diaz LA, Shiu KK, Kim TW, et al. Pembrolizumab versus chemotherapy for microsatellite instability-high or mismatch repair-deficient metastatic colorectal cancer (KEYNOTE-177): final analysis of a randomised, open-label, phase 3 study. *Lancet Oncol*. 2022;23(5):659-670.
13. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773-1784.
14. Stein-O'Brien GL, Le DT, Jaffee EM, Fertig EJ, Zaidi N. Converging on a cure: the roads to predictive immunotherapy. *Cancer Discov*. 2023;13:1053-1057. Published online April 17, 2023:OF1-OF5.

15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
16. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. Published online October 22. 2020. http://arxiv.org/abs/2010.11929.
17. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion*. 2020;58(October 2019):82-115.
18. Osband I, Blundell C, Pritzel A, van Roy B. Deep exploration via bootstrapped DQN. *Adv Neural Inf Process Syst*. 2016;29.
19. Flügge T, Gaudin R, Sabatakakis A, et al. Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. *Sci Rep*. 2023;13(1):2296.
20. Li TZ, Xu K, Gao R, et al. Time-distance vision transformers in lung cancer diagnosis from longitudinal computed tomography. *Proc SPIE Int Soc Opt Eng*. 2023:12464.
21. Deininger L, Stimpel B, Yuce A, et al. A comparative study between vision transformers and CNNs in digital pathology. *arXiv preprint arXiv*. 2022;2206:00389.
22. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
23. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst Appl*. 1998;13(4):18-28.
24. Ilse M, Jakub T, Max W. *"Attention-based deep multiple instance learning." International conference on machine learning*. PMLR; 2018. p. 2127-2136.
25. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. The International Conference on Learning Representations (ICLR). Published online September 9. 2016. http://arxiv.org/abs/1609.02907.
26. Cox DR. Regression models and life-tables. *J R Stat Soc*. 1972;34(1-2):187-220.
27. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. *ACM Comput Surv*. 2019;51:110.
28. Pölsterl S. Scikit-survival: a library for time-to-event analysis built on top of scikit-learn. *J Mach Learn Res*. 2020;21:1-6.
29. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841-860.
30. Kundu S. AI in medicine must be explainable. *Nat Med*. 2021;27(8):1328. 1328.
31. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci*. 2019;116(44):22071-22080.
32. Kadir T, Brady M. Saliency, scale and image description. *Int J Comput Vis*. 2001;45(2):83-105.
33. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *Br Med J*. 2021:372.
34. Chowell D, Yoo SK, Valero C, et al. Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nat Biotechnol*. 2022;40(4):499-506.
35. Kong JH, Ha D, Lee J, et al. Network-based machine learning approach to predict immunotherapy response in cancer patients. *Nat Commun*. 2022;13(1):3703.
36. Peng J, Xiao L, Zou D, Han L. A somatic mutation signature predicts the best overall response to anti-programmed cell death protein-1 treatment in epidermal growth factor receptor/anaplastic lymphoma kinase-negative non-squamous non-small cell lung cancer. *Front Med (Lausanne)*. 2022;9:808378.
37. Fang C, Xu D, Su J, Dry JR, Linghu B. DeePaN: deep patient graph convolutional network integrating clinico-genomic evidence to stratify lung cancers for immunotherapy. *NPJ Digit Med*. 2021;4(1):14.
38. Ahn BC, So JW, Synn CB, et al. Clinical decision support algorithm based on machine learning to assess the clinical response to anti—programmed death-1 therapy in patients with non—small-cell lung cancer. *Eur J Cancer*. 2021;153:179-189.
39. Wang Y, Chen L, Ju L, Xiao Y, Wang X. Tumor mutational burden related classifier is predictive of response to PD-L1 blockade in locally advanced and metastatic urothelial carcinoma. *Int Immunopharmacol*. 2020;87:106818.
40. Peng J, Zou D, Gong W, Kang S, Han L. Deep neural network classification based on somatic mutations potentially predicts clinical benefit of immune checkpoint blockade in lung adenocarcinoma. *Oncoimmunology*. 2020;9(1):1734156.
41. Xie F, Zhang J, Wang J, et al. Multifactorial deep learning reveals pan-cancer genomic tumor clusters with distinct immunogenomic landscape and response to immunotherapy. *Clin Cancer Res*. 2020;26(12):2908-2920.
42. Wang L, Zhang H, Pan C, et al. Predicting durable responses to immune checkpoint inhibitors in non-small-cell lung cancer using a multi-feature model. *Front Immunol*. 2022;13:829634.
43. Wiesweg M, Mairinger F, Reis H, et al. Machine learning reveals a PD-L1—independent prediction of response to immunotherapy of non-small cell lung cancer by gene expression context. *Eur J Cancer*. 2020;140(September):76-85.
44. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248-262.
45. Ren C, Li J, Zhou Y, Zhang S, Wang Q. Typical tumor immune microenvironment status determine prognosis in lung adenocarcinoma. *Transl Oncol*. 2022;18:101367.
46. Xu L, Jian X, Liu Z, et al. Construction and validation of an immune cell signature score to evaluate prognosis and therapeutic efficacy in hepatocellular carcinoma. *Front Genet*. 2021;12:741226.
47. Chen R, Wu W, Liu T, et al. Large-scale bulk RNA-seq analysis defines immune evasion mechanism related to mast cell in gliomas. *Front Immunol*. 2022;13:914001.
48. Zhang Y, Goto Y, Yagishita S, et al. Machine learning-based exceptional response prediction of nivolumab monotherapy with circulating microRNAs in non-small cell lung cancer. *Lung Cancer*. 2022;173:107-115.
49. Ma N, Li J, Lv L, Li C, Li K, Wang B. Bioinformatics evaluation of a novel angiogenesis related genes-based signature for predicting prognosis and therapeutic efficacy in patients with gastric cancer. *Am J Transl Res*. 2022;14:4532-4548.
50. Zhang Z, Wang ZX, Chen YX, et al. Integrated analysis of single-cell and bulk RNA sequencing data reveals a pan-cancer stemness signature predicting immunotherapy response. *Genome Med*. 2022;14(1):45.
51. Liu Z, Xu H, Weng S, Ren Y, Han X. Stemness refines the classification of colorectal cancer with stratified prognosis, multi-omics landscape, potential mechanisms, and treatment options. *Front Immunol*. 2022;13:828330.
52. Chen Y, Miao S, Zhao W. Identification and validation of significant gene mutations to predict clinical benefit of immune checkpoint inhibitors in lung adenocarcinoma. *Am J Transl Res*. 2021;13(3):1051-1063.
53. Li G, Luo Q, Wang X, Zeng F, Feng G, Che G. Deep learning reveals cuproptosis features assist in predict prognosis and guide immunotherapy in lung adenocarcinoma. *Front Endocrinol (Lausanne)*. 2022;13:970269.
54. Prelaj A, Boeri M, Robuschi A, et al. Machine learning using real-world and translational data to improve treatment selection for NSCLC patients treated with immunotherapy. *Cancers (Basel)*. 2022;14(2):435.
55. Zheng K, Gao L, Hao J, Zou X, Hu X. An immunotherapy response prediction model derived from proliferative CD4+ T cells and antigen-presenting monocytes in ccRCC. *Front Immunol*. 2022;13:972227.
56. Huang Y, Liu H, Liu X, et al. The chemokines initiating and maintaining immune hot phenotype are prognostic in ICB of HNSCC. *Front Genet*. 2022;13:820065.
57. Vathiotis IA, Yang Z, Reeves J, et al. Models that combine transcriptomic with spatial protein information exceed the predictive value for either single modality. *NPJ Precis Oncol*. 2021;5(1):45.
58. Xu H, Liu Z, Weng S, et al. Artificial intelligence-driven consensus gene signatures for improving bladder cancer clinical outcomes identified by multi-center integration analysis. *Mol Oncol*. 2022;16:4023-4042. Published online, . Accessed September 22, 2022.

59. Kim K, Kim HS, Kim JY, et al. Predicting clinical benefit of immunotherapy by antigenic or functional mutations affecting tumour immunogenicity. *Nat Commun*. 2020;11(1):951.
60. Liu Z, Xu H, Weng S, et al. Machine learning algorithm-generated and multi-center validated melanoma prognostic signature with inspiration for treatment management. *Cancer Immunol Immunother*. 2023;72:599-615. Published online 2022.
61. Hu X, Wu L, Yao Y, et al. The integrated landscape of eRNA in gastric cancer reveals distinct immune subtypes with prognostic and therapeutic relevance. *iScience*. 2022;25(10):105075.
62. Wang Z, Wang Y, Yang T, et al. Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients. *Brief Bioinform*. 2021;22(5):bbab032.
63. Zhang Z, Chen L, Chen H, et al. Pan-cancer landscape of T-cell exhaustion heterogeneity within the tumor microenvironment revealed a progressive roadmap of hierarchical dysfunction associated with prognosis and therapeutic efficacy. *EBioMedicine*. 2022;83:104207.
64. Liu R, Dollinger E, Nie Q. Machine learning of single cell transcriptomic data from anti-PD-1 responders and non-responders reveals distinct resistance mechanisms in skin cancers and PDAC. *Front Genet*. 2022;12:806457.
65. Zhang N, Zhang H, Wu W, et al. Machine learning-based identification of tumor-infiltrating immune cell-associated lncRNAs for improving outcomes and immunotherapy responses in patients with low-grade glioma. *Theranostics*. 2022;12(13):5931-5948.
66. Filipski K, Scherer M, Zeiner KN, et al. DNA methylation-based prediction of response to immune checkpoint inhibition in metastatic melanoma. *J Immunother Cancer*. 2021;9(7):e002226.
67. Xu B, Lu M, Yan L, et al. A pan-cancer analysis of predictive methylation signatures of response to cancer immunotherapy. *Front Immunol*. 2021;12:796647.
68. Zhou M, Zhang Z, Bao S, et al. Computational recognition of lncRNA signature of tumor-infiltrating B lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer. *Brief Bioinform*. 2021;22(3):bbaa047.
69. Liu Z, Liu L, Weng S, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. 2022;13(1):816.
70. Pan X, Zhang C, Wang J, et al. Epigenome signature as an immunophenotype indicator prompts durable clinical immunotherapy benefits in lung adenocarcinoma. *Brief Bioinform*. 2022;23(1):bbab481.
71. The Cancer Genome Atlas Program (TCGA), Institute, National Cancer. Available at https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga. Accessed November 20, 2023.
72. NCBI. Gene Expression Omnibus. Available at https://www.ncbi.nlm.nih.gov/geo/. Accessed February 3, 2023.
73. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.
74. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal complementary data sources and analysis options. *Sci Signal*. 2013;6(269):1-20.
75. Subbiah V, Solit DB, Chan TA, Kurzrock R. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) $\geq$10: a decision centered on empowering patients and their physicians. *Ann Oncol*. 2020;31(9):1115-1118.
76. lászló Barabási A, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12:56-68.
77. Wu CC, Wang YA, Livingston JA, Zhang J, Futreal PA. Prediction of biomarkers and therapeutic combinations for anti-PD-1 immunotherapy using the global gene network association. *Nat Commun*. 2022;13(1):42.
78. Dercle L, Fronheiser M, Lu L, et al. Identification of non-small cell lung cancer sensitive to systemic cancer therapies using radiomics. *Clin Cancer Res*. 2020;26(9):2151-2162.
79. He B, Dong D, She Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer*. 2020;8(2):1-10.
80. Khorrami M, Prasanna P, Gupta A, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non—small cell lung cancer. *Cancer Immunol Res*. 2020;8(1):108-119.
81. Mu W, Jiang L, Shi Y, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer*. 2021;9(6):1-15.
82. Tian P, He B, Mu W, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics*. 2021;11(5):2098-2107.
83. Tunali I, Gray JE, Qi J, et al. Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: an early report. *Lung Cancer*. 2019;129(January):75-79.
84. Tunali I, Tan Y, Gray JE, et al. Hypoxia-related radiomics and immunotherapy response: a multicohort study of non-small cell lung cancer. *JNCI Cancer Spectr*. 2021;5(4):1-11.
85. Vaidya P, Bera K, Patil PD, et al. Novel, non-invasive imaging approach to identify patients with advanced non-small cell lung cancer at risk of hyperprogressive disease with immune checkpoint blockade. *J Immunother Cancer*. 2020;8(2):1-11.
86. Deng K, Wang L, Liu Y, et al. A deep learning-based system for survival benefit prediction of tyrosine kinase inhibitors and immune checkpoint inhibitors in stage IV non-small cell lung cancer patients: a multicenter, prognostic study. *EClinicalMedicine*. 2022;51:101541.
87. Gong J, Bao X, Wang T, et al. A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. *Oncoimmunology*. 2022;11(1):2028962.
88. He BX, Zhong YF, Zhu YB, et al. Deep learning for predicting immunotherapeutic efficacy in advanced non-small cell lung cancer patients: a retrospective study combining progression-free survival risk and overall survival risk. *Transl Lung Cancer Res*. 2022;11(4):670-685.
89. Ren Q, Xiong F, Zhu P, et al. Assessing the robustness of radiomics/deep learning approach in the identification of efficacy of anti—PD-1 treatment in advanced or metastatic non-small cell lung carcinoma patients. *Front Oncol*. 2022;12:952749.
90. Dercle L, Zhao B, Gönen M, et al. Early readout on overall survival of patients with melanoma treated with immunotherapy using a novel imaging analysis. *JAMA Oncol*. 2022;8(3):385-392.
91. Basler L, Gabryś HS, Hogan SA, et al. Radiomics, tumor volume, and blood biomarkers for early prediction of pseudoprogression in patients with metastatic melanoma treated with immune checkpoint inhibition. *Clin Cancer Res*. 2020;26(16):4414-4425.
92. Brendlin AS, Peisen F, Almansour H, et al. A machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. *J Immunother Cancer*. 2021;9(11):e003261.
93. Peisen F, Hänsch A, Hering A, et al. Combination of whole-body baseline CT radiomics and clinical parameters to predict response and survival in a stage-IV melanoma cohort undergoing immunotherapy. *Cancers (Basel)*. 2022;14(12):2992.
94. George E, Flagg E, Chang K, et al. Radiomics-based machine learning for outcome prediction in a multicenter phase II study of programmed death-ligand 1 inhibition immunotherapy for glioblastoma. *Am J Neuroradiol*. 2022;43(5):675-681.
95. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol*. 2019;30(6):998-1004.
96. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol*. 2018;19(9):1180-1191.
97. Ligero M, Garcia-Ruiz A, Viaplana C, et al. A CT-based radiomics signature is associated with response to immune checkpoint inhibitors in advanced solid tumors. *Radiology*. 2021;299(1):109-119.

98. Harder N, Schönmeyer R, Nekolla K, et al. Automatic discovery of image-based signatures for ipilimumab response prediction in malignant melanoma. *Sci Rep*. 2019;9(1):7449.
99. Park S, Ock CY, Kim H, et al. Artificial intelligence—powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non—small-cell lung cancer. *J Clin Oncol*. 2022;40(17):1916-1928.
100. Baxi V, Lee G, Duan C, et al. Association of artificial intelligence-powered and manual quantification of programmed death-ligand 1 (PD-L1) expression with outcomes in patients treated with nivolumab ± ipilimumab. *Mod Pathol*. 2022;35(11):1529-1539.
101. Althammer S, Tan TH, Spitzmüller A, et al. Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. *J Immunother Cancer*. 2019;7(1):1-12.
102. Johannet P, Coudray N, Donnelly DM, et al. Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. *Clin Cancer Res*. 2021;27(1):131-140.
103. Hu J, Cui C, Yang W, et al. Using deep learning to predict anti-PD-1 response in melanoma and lung cancer patients from histopathology images. *Transl Oncol*. 2021;14(1):100921.
104. Chen Y, Jia K, Sun Y, et al. Predicting response to immunotherapy in gastric cancer via multi-dimensional analyses of the tumour immune microenvironment. *Nat Commun*. 2022;13(1):4851.
105. Choi S, Cho SI, Ma M, et al. Artificial intelligence—powered programmed death ligand 1 analyser reduces interobserver variation in tumour proportion score for non—small cell lung cancer with better prediction of immunotherapy response. *Eur J Cancer*. 2022;170:17-26.
106. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer*. 2020;1(8):800-810.
107. Kather JN, Heij LR, Grabsch HI, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer*. 2020;1(8):789-799.
108. Benzekry S, Grangeon M, Karlsen M, et al. Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data. *Cancers (Basel)*. 2021;13(24):6210.
109. Hellwig K, Ellmann S, Eckstein M, et al. Predictive value of multiparametric MRI for response to single-cycle induction chemoimmunotherapy in locally advanced head and neck squamous cell carcinoma. *Front Oncol*. 2021;11:734872.
110. Weber JS, Sznol M, Sullivan RJ, et al. A serum protein signature associated with outcome after anti—PD-1 therapy in metastatic melanoma. *Cancer Immunol Res*. 2018;6(1):79-86.
111. Lui TKL, Cheung KS, Leung WK. Machine learning models in the prediction of 1-year mortality in patients with advanced hepatocellular cancer on immunotherapy: a proof-of-concept study. *Hepatol Int*. 2022;16(4):879-891.
112. Arbour KC, Luu AT, Luo J, et al. Deep learning to estimate RECIST in patients with NSCLC treated with PD-1 blockade. *Cancer Discov*. 2021;11(1):59-67.
113. Wu Y, Zhu W, Wang J, et al. Using machine learning for mortality prediction and risk stratification in atezolizumab-treated cancer patients: integrative analysis of eight clinical trials. *Cancer Med*. 2023;12:3744-3757. Published online July 24, 2022.
114. Li C, Tian C, Zeng Y, et al. Machine learning and bioinformatics analysis revealed classification and potential treatment strategy in stage 3—4 NSCLC patients. *BMC Med Genomics*. 2022;15(1):33.
115. Liang H, Jo JH, Zhang Z, et al. Predicting cancer immunotherapy response from gut microbiomes using machine learning models. *Oncotarget*. 2022;13:876-889.
116. Muller M, Hummelink K, Hurkmans DP, et al. A serum protein classifier identifying patients with advanced non—small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors. *Clin Cancer Res*. 2020;26(19):5188-5197.
117. Iivanainen S, Ekström J, Virtanen H, Kataja VV, Koivunen JP. Predicting objective response rate (ORR) in immune checkpoint inhibitor (ICI) therapies with machine learning (ML) by combining clinical and patient-reported data. *Appl Sci*. 2022;12(3):1563.
118. Wei C, Wang M, Gao Q, et al. Dynamic peripheral blood immune cell markers for predicting the response of patients with metastatic cancer to immune checkpoint inhibitors. *Cancer Immunol Immunother*. 2023;72(1):23-37.
119. Sun S, Guo W, Lv F, et al. Comprehensive analysis of ferroptosis regulators in lung adenocarcinomas identifies prognostic and immunotherapy-related biomarkers. *Front Mol Biosci*. 2021;8:587436.
120. Abuhelwa AY, Kichenadasse G, McKinnon RA, Rowland A, Hopkins AM, Sorich MJ. Machine learning for prediction of survival outcomes with immune-checkpoint inhibitors in urothelial cancer. *Cancers (Basel)*. 2021;13(9):2001.
121. Gupta A, Arora P, Brenner D, et al. Risk prediction using Bayesian networks: an immunotherapy case study in patients with metastatic renal cell carcinoma. *JCO Clin Cancer Inform*. 2021;5:326-337.
122. Sidhom JW, Oliveira G, Ross-Macdonald P, et al. Deep learning reveals predictive sequence concepts within immune repertoires to immunotherapy. *Sci Adv*. 2022;8:eabq5089.
123. Rounis K, Makrakis D, Papadaki C, et al. Prediction of outcome in patients with non-small cell lung cancer treated with second line PD-1/PDL-1 inhibitors based on clinical parameters: results from a prospective, single institution study. *PLoS One*. 2021;16(6):e0252537.
124. Madonna G, Masucci GV, Capone M, et al. Clinical categorization algorithm (CLICAL) and machine learning approach (SRF-CLICAL) to predict clinical benefit to immunotherapy in metastatic melanoma patients: real-world evidence from the Istituto Nazionale Tumori IRCCS Fondazione Pascale, Napoli, It. *Cancers (Basel)*. 2021;13(16):4164.
125. Liu X, Xiao Z, Song Y, Zhang R, Li X, Du Z. A machine learning-aided framework to predict outcomes of anti-PD-1 therapy for patients with gynecological cancer on incomplete post-marketing surveillance dataset. *IEEE Access*. 2021;9:120464-120480.
126. Wu H, Zheng X, Pan T, et al. Dynamic microbiome and metabolome analyses reveal the interaction between gut microbiota and anti-PD-1 based immunotherapy in hepatocellular carcinoma. *Int J Cancer*. 2022;151(8):1321-1334.
127. Harel M, Lahav C, Jacob E, et al. Longitudinal plasma proteomic profiling of patients with non-small cell lung cancer undergoing immune checkpoint blockade. *J Immunother Cancer*. 2022;10(6):e004582.
128. Ko JJ, Xie W, Kroeger N, et al. The International Metastatic Renal Cell Carcinoma Database Consortium model as a prognostic tool in patients with metastatic renal cell carcinoma previously treated with first-line targeted therapy: a population-based study. *Lancet Oncol*. 2015;16(3):293-300.
129. 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.
130. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer*. 2022;22(2):114-126.
131. Yang Y, Yang J, Shen L, et al. A multi-omics-based serial deep learning approach to predict clinical outcomes of single-agent anti-PD-1/PD-L1 immunotherapy in advanced stage non-small-cell lung cancer. *Am J Transl Res*. 2021;13(2):743-756.
132. Shen R, Postow MA, Adamow M, et al. LAG-3 expression on peripheral blood cells identifies patients with poorer outcomes after immune checkpoint blockade. *Sci Transl Med*. 2021;13:eabf5107.
133. Park C, Na KJ, Choi H, et al. Tumor immune profiles noninvasively estimated by FDG PET with deep learning correlate with immunotherapy response in lung adenocarcinoma. *Theranostics*. 2020;10(23):10838-10848.
134. Mu W, Jiang L, Zhang JY, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun*. 2020;11(1):5228.
135. Lipkova J, Chen RJ, Chen B, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*. 2022;40(10):1095-1110.
136. The Cancer Imaging Archive. National Cancer Institute. Available at https://www.cancerimagingarchive.net/. Accessed March 3, 2023.
137. Elkin R, Oh JH, Liu YL, et al. Geometric network analysis provides prognostic information in patients with high grade serous carcinoma

of the ovary treated with immune checkpoint inhibitors. *NPJ Genom Med*. 2021;6(1):99.
138. Chen Q, Zhang L, Mo X, et al. Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging*. 2021;49(1):345-360.
139. Zhang C, de A F Fonseca L, Shi Z, et al. Systematic review of radiomic biomarkers for predicting immune checkpoint inhibitor treatment outcomes. *Methods*. 2021;188(2021):61-72.
140. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. 2019;25(7):1054-1056.
141. Shamai G, Livne A, Polónia A, et al. Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. *Nat Commun*. 2022;13(1):6753.
142. Zeng Q, Klein C, Caruso S, et al. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *J Hepatol*. 2022;77(1):116-127.
143. Prelaj A, Galli EG, Miskovic V, et al. Real-world data to build explainable trustworthy artificial intelligence models for prediction of immunotherapy efficacy in NSCLC patients. *Front Oncol*. 2023;12:1-14.
144. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf Fusion*. 2019;50:71-91.
145. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206-215.
146. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3(11):e745-e750.
147. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374.
148. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351-1363.
149. Topol EJ. Welcoming new guidelines for AI clinical research. *Nat Med*. 2020;26(9):1318-1320.
150. Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health*. 2022;4(4):e214-e215.