

REVIEW

Artificial Intelligence & Precision Oncology

# Radiomics in Antineoplastic Agents Development: Application and Challenge in Response Evaluation

Jiazheng Li, Lei Tang\*

Department of Radiology, Peking University Cancer Hospital & Institute,  
Key Laboratory of Carcinogenesis and Translational Research  
(Ministry of Education), Beijing 100142, China

**Key words:** radiomics; deep learning; machine learning; antineoplastic agents; response evaluation

**Abstract** The recent spring up of the antineoplastic agents and the prolonged survival bring both challenge and chance to radiological practice. Radiological methods including CT, MRI and PET play an increasingly important role in evaluating the efficacy of these antineoplastic drugs. However, different antineoplastic agents potentially induce different radiological signs, making it a challenge for radiological response evaluation, which depends mainly on one-sided morphological response evaluation criteria in solid tumors (RECIST) in the status quo of clinical practice. This brings opportunities for the development of radiomics, which is promising to serve as a surrogate for response evaluations of anti-tumor treatments. In this article, we introduce the basic concepts of radiomics, review the state-of-art radiomics researches with highlights of radiomics application in predictions of molecular biomarkers, treatment response, and prognosis. We also provide in-depth analyses on major obstacles and future direction of this new technique in clinical investigations on new antineoplastic agents.

## OVERVIEW OF IMAGING RESPONSE EVALUATION

**T**HE latest few decades have witnessed great advances in antineoplastic agents, which mainly focus on chemotherapy, immunotherapy, and molecular targeted therapy.<sup>[1-3]</sup> Following the spring up of the antineoplastic agents and the consequential prolonged survival, traditional gold standard for response evaluation, *i.e.*, the endpoint

of overall survival (OS), has met the Achilles heel. As an important alternative endpoint, the application of radiological methods, such as CT, MRI, and PET, have experienced unprecedented explosive growth in oncological clinical trials.<sup>[4]</sup> In order to standardize the imaging endpoint processes, the Center for Drug Evaluation (CDE) of National Medical Products Administration of China published *The Principles and Procedures of Clinical Trial Imaging Endpoint Process Standards for Antineoplastic Agents* on January 15, 2021.<sup>[4]</sup>

However, there still exist several limitations and hurdles in clinical practice. Antineoplastic agents with different biological mechanisms and administered in different treatment lines may present different radiological signs regarding response to the treatment. The Response Evaluation Criteria in Solid Tumors (RECIST) simply measures changes in long diameter of targeted

Received August 19, 2021; accepted September 14, 2021; published online September 21, 2021.

\*Corresponding author E-mail: terrytang78@163.com

This work is supported by the Beijing Natural Science Foundation (Z180001; Z200015), PKU-Baidu Fund (2020BD027), and the National Natural Science Foundation of China (91959205).

lesions based on morphology, and appears no longer meet the needs of evaluating the existing anticancer treatments.<sup>[5]</sup> For instance, good response to Imatinib, the first-line targeted treatment agent for gastrointestinal stromal tumor (GIST), may occur tumor necrosis, cystic or myxoid degeneration, resulting in decreased density on CT images, whereas tumor size may maintain unchanged, or even enlarged caused by degeneration of tumorous tissues, which might be determined as stable or progression of the disease according to RECIST.<sup>[5]</sup>

Moreover, ongoing evolution of tumor may happen and cause intratumoral heterogeneity. According to our clinical observation on a cohort of GIST patients treated with ripretinib, intratumoral heterogeneity (manifested as spatial heterogeneity and/or temporal heterogeneity) happened in as high as 35% of the cases. Such heterogeneity brings challenges in radiological evaluation, and innovative predictors are urgently needed to accurately evaluate response of tumor to the treatments of antineoplastic agents.

## BASIC CONCEPTS IN RADIOMICS

Under the concept of precision medicine, radiomics has been recently introduced into radiological practices to help management of huge volume of data, and has shown potentials in integrating multi-modality radiological data to facilitate tumor evaluation.<sup>[6]</sup> This data-mining driven method operates on the hypothesis that radiological images contain huge amount of information of tumor biological behavior which cannot be interpreted directly by human subjective consciousness, but can be extracted through texture analysis and analyzed with integrating models by computer.<sup>[7]</sup>

Some classical machine-learning algorithms are available in the field of radiomics, such as random forest, logistic regression model, and support vector machine.<sup>[8, 9]</sup> By using these methods, quantitative descriptors based on shape, size, and texture can be extracted from region of interests defined by radiologists.<sup>[7]</sup> After reducing the dimensionality of these features, a classifier model could be built through machine learning method, and decision-supporting could be achieved based on various practical scenarios and thereafter meet different requirements for radiology.<sup>[7]</sup> As a state-of-art subfield of artificial intelligence (AI), deep learning could be adopted in imaging procedures. Unlike machine-learning algorithms

that has designed features by human, deep learning can automatically understand and manipulate raw and large datasets, *i.e.*, the pixel values of an image, select features needed for pattern recognition via iterative adjustment, and finally distinguish the raw dataset inputs.<sup>[10]</sup> Convolutional neural network (CNN) is the commonly used deep learning algorithm in the studies of radiomics.<sup>[11, 12]</sup>

## RESEARCHES OF RADIOMICS IN RESPONSE EVALUATION

Facilitated with AI, radiology could comprehensively evaluate the overall intratumoral characteristics compared to the regional representative nature of biopsy pathology.<sup>[13]</sup> Currently, there is a growing interest of radiomics applications in antineoplastic agents development, including prediction of molecular biomarker of tumor, prediction of response to therapy, and predicting patient outcomes. Different algorithms have been tested and some promising application fields were proposed.<sup>[9, 14-32]</sup> The most representative studies are presented in **Table 1**.

### Prediction of molecular biomarker

There were several studies exploring radiomic parameters that correlate with molecular biomarker of tumor, which may signify the response of the tumor to anti-cancer therapies. PD-1/PD-L1 signaling can be blocked by immune checkpoint inhibitors to obtain an enhanced antitumor immune activity.<sup>[3]</sup> A model to predict PD-L1 expression in non-small cell lung carcinoma (NSCLC) was built by CNN, which showed high predictive ability with an area under the ROC curve (AUC) of 0.78, and it also could predict efficacy of the immunotherapy.<sup>[20]</sup> Trastuzumab, a humanized monoclonal antibody that targets the overexpression of HER2 on tumor cells to promote their lysis, is used in targeted therapy for HER2-positive gastric cancer patients.<sup>[33]</sup> Li *et al.* investigated the value of CT radiomics in predicting HER2 status of gastric cancer by least absolute shrinkage and selection operator (LASSO) logistic regression model, and got a high accuracy (AUC = 0.771), which indicates its value in guiding the screening of patients who are susceptible to targeted therapy.<sup>[9]</sup> GIST with mutations other than KIT exon 11 mutant usually respond poorly to Imatinib.<sup>[1]</sup> Xu *et al.* reported a radiogenomics study on GIST. They built a model with radiomic features, anatomic locations, and CD34

stain levels integrated together into a support vector machine algorithm to differentiate KIT exon 11 mutation positive from negative, and demonstrated a high accuracy (AUC=0.904).<sup>[14]</sup>

Solid tumors consist of not only neoplastic cells, but also a variety types of nonmalignant cells, such as stromal and immune cells, which interact with cancer cells and play a part in its biological behavior.<sup>[34, 35]</sup> Tumor stroma status is positively correlated with epithelial-mesenchymal transition (EMT) and angiogenesis pathways, and is related to the upregulation of cytokines of immunosuppression function.<sup>[34]</sup> Most recently, a deep residual neural network with the deformable image augmentation by attention U-net was developed to detect gastric cancer stroma status determined by immunohistochemistry of  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA) and periostin (POSTN). The diagnostic accuracy of this deep-learning model was 0.88-0.89 with a high level of repeatability and stability.<sup>[17]</sup>

There were also some studies focused on immunotherapy. Immune checkpoint inhibitors act by reinvigorating preexisting antitumor T-cell responses and are most effective in tumors with high density of CD8+ T-cell.<sup>[35]</sup> In a multicenter and multi-cohort study, a regression method was explored to build a linear elastic-net model for detecting the abundance of CD8 cell quantified by RNA-seq data in pan-cancer.<sup>[15]</sup> This machine learning model was shown to be highly effective in classifying high versus low abundance of CD8 infiltrate (AUC = 0.74) and identifying patients who might respond well to anti-PD-1 or anti-PD-L1 therapy (P = 0.013) at the baseline.<sup>[15]</sup> Current studies suggest that radiomic features could evaluate lesion longitudinally and noninvasively, which overcome the temporal heterogeneity of tumors and provide valuable information to clinicians for tailored treatment strategy.

### Prediction of response to therapy and patient outcomes

The current generation of targeted and immunotherapies can produce a clear clinical benefit, which is poorly captured by RECIST.<sup>[36]</sup> Studies have shown that radiomics has potential to overcome this challenge.

In a multicenter study, machine learning based radiomics has been explored to predict the added benefit of adjuvant chemotherapy in early-stage NSCLC.<sup>[23]</sup> In combination with T stage, N stage, and lymphovascular invasion, the machine learning model was highly effective in selecting patients who would have

a significantly prolonged survival while receiving adjuvant chemotherapy ( $p=0.03$ ). Radiomic features also have correlation with the response to immunotherapy. Trebeschi S. *et al.* reported that in NSCLC, lesions responded to immunotherapy presented with more heterogeneous morphological features; they developed a model with random forest algorithm to predict progression at lesion level and the OS at individual patient level before therapy, which was demonstrated with AUC of 0.70-0.83, and of 0.76, respectively.<sup>[22]</sup>

Some novel response patterns, such as pseudoprogression induced by immune checkpoint inhibition (ICI), have brought greater challenge for the response evaluation base on RECIST criteria.<sup>[36]</sup> A study declared that during ICI treatment on melanoma, changes in radiomic features from homogeneous to heterogeneous indicate a lower chance of pseudoprogression, and the model built on the relative difference of radiomic features between baseline and 3-months post treatment in combination with blood LDH level of 3-months post-treatment achieved the highest accuracy (AUC=0.82; sensitivity=0.81; specificity=0.73) in early differentiation of pseudoprogression from true progress.<sup>[27]</sup>

Studies investigating the predictive value of radiomics in the treatment of molecular targeted therapy have shown promising results. To predict the response and survival of metastatic colorectal cancer patients treated with agents targeting EGFR pathway, the change in imaging features on CT scans between baseline and the 8-week post-treatment were used to build a model by random forest algorithm, which was shown to have a satisfactory performance (AUC=0.80), outperformed that using KRAS mutational status at baseline (AUC=0.67) and that using tumor shrinkage 8-week after the treatment based on RECIST criteria (AUC=0.75) for detection of treatment sensitivity, and was strongly associated with the OS.<sup>[25]</sup>

### CHALLENGES AND POTENTIAL SOLUTIONS

The quality and completeness of published radiomics studies are variable. It has been recognized that some radiomics studies are inadequate to conduct high-quality results. One of the major obstacle is the relatively small sample size, especially when new treatment schemes are involved. Among radiomics studies, the average sample size is 232 patients,

**Table 1.** Representative radiomics studies regarding evaluation or prediction of tumor response to anti-neoplastic therapies

| Author                     | Journal             | Publication Year | Study Type                  | Cancer            | Sample Size |            | Image Type                        | Orientation                                                          | Segmentation                                                      | Radiomics Features                                                                   | Algorithms                                                        |
|----------------------------|---------------------|------------------|-----------------------------|-------------------|-------------|------------|-----------------------------------|----------------------------------------------------------------------|-------------------------------------------------------------------|--------------------------------------------------------------------------------------|-------------------------------------------------------------------|
|                            |                     |                  |                             |                   | Training    | Validation |                                   |                                                                      |                                                                   |                                                                                      |                                                                   |
| Sun R <sup>[15]</sup>      | Lancet Oncol        | 2018             | Retro-spective; Multicenter | Pan-cancer        | 135         | 356        | CE-CT                             | Estimate tumor CD8 cell count                                        | Primary lesion, biggest lesion, target lesions; a peripheral ring | First-, second-order features, and volumes features                                  | linear elastic-net model                                          |
| Jiang Y <sup>[16]</sup>    | Ann On-col          | 2020             | Retro-spective; Multicenter | Gastric cancer    | 262         | 1516       | Abdomen portal venous phase CE-CT | Predict tumor immune status                                          | Primary tumor; a peripheral ring                                  | First-, second-, and higher order features and shape features                        | LASSO method                                                      |
| Jiang Y <sup>[17]</sup>    | Lancet Digit Health | 2021             | Retro-spective; Multicenter | Gastric cancer    | 321         | 1888       | Abdomen portal venous phase CE-CT | Assess tumor stroma microenvironment                                 | primary tumor                                                     | Deep learning features                                                               | Attention U-net for deformable image augmentation and deep ResNet |
| Mu W <sup>[18]</sup>       | Nat Commun          | 2020             | Retro-spective; Multicenter | NSCLC             | 429         | 468        | <sup>18</sup> F-FDG PET/CT        | Define EGFR mutation status                                          | Primary tumor; a peripheral ring                                  | Deep learning features                                                               | 2D small-residual-convolutional-network                           |
| Rossi G <sup>[19]</sup>    | Cancer Res          | 2021             | Retro-spective; Multicenter | NSCLC             | 109         | 61         | Chest CT without contrast         | Define EGFR mutation status                                          | Tumor corresponding to the biological sample                      | Shape and textural features                                                          | Support Vector Machine                                            |
| Tian P <sup>[20]</sup>     | Theranostics        | 2021             | Retrospective               | NSCLC             | 750         | 283        | CT                                | Predict PD-L1 expression                                             | Primary tumor                                                     | Deep learning features                                                               | Deep learning network                                             |
| Shi R <sup>[21]</sup>      | Am J Cancer Res     | 2020             | Retro-spective; Multicenter | Colorectal cancer | 124         | 35         | Abdomen portal venous phase CE-CT | Distinguish patients with RAS or BRAF mutation                       | Metastatic liver lesions                                          | First-order features, shape-based features, and textural features                    | ANN method                                                        |
| Trebesch S <sup>[22]</sup> | Ann On-col          | 2019             | Retro-spective; Multicenter | NSCLC; Melanoma   | 203         | 301        | Chest CT or chest-abdomen CE-CT   | Predict lesion's progression and patient's OS before immunotherapy   | Lesions with diameter > 5 mm                                      | Laplacian of Gaussian filters, wavelets decompositions, and non-linearities features | Random forest                                                     |
| Vaidya P <sup>[23]</sup>   | Lancet Digit Health | 2020             | Retro-spective; Multicenter | NSCLC             | 329         | 196        | Chest CT with or without contrast | Predict survival after adjuvant chemotherapy                         | Target nodule; a peripheral ring                                  | Gabor, Haralick, Laws, Laplace, and Collage feature                                  | LASSO method                                                      |
| Vaidya P <sup>[24]</sup>   | J Immunother Cancer | 2020             | Retrospective               | NSCLC             | 30          | 79         | CT with or without contrast       | Distinguish hyperprogression from response patterns to immunotherapy | All target lesions; a peripheral ring                             | Gabor, Haralick, Laws, Laplace, and Collage feature                                  | Random forest                                                     |
| Dercle L <sup>[25]</sup>   | J Natl Cancer Inst  | 2020             | Retro-spective; Multicenter | Colorectal cancer | 439         | 229        | Abdomen portal venous phase CE-CT | Predict tumor sensitivity to anti-EGFR treatment                     | Metastatic liver lesions                                          | Radiomics features*, deep learning features                                          | Random forest                                                     |

**Continued Table 1.** Representative radiomics studies regarding evaluation or prediction of tumor response to anti-neoplastic therapies

| Author                   | Journal         | Publication Year | Study Type                 | Cancer         | Sample Size |                  | Image Type                        | Orientation                                                               | Segmentation                                       | Radiomics Features                                                                                                                 | Algorithms                                                      |
|--------------------------|-----------------|------------------|----------------------------|----------------|-------------|------------------|-----------------------------------|---------------------------------------------------------------------------|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
|                          |                 |                  |                            |                | Training    | Validation       |                                   |                                                                           |                                                    |                                                                                                                                    |                                                                 |
| Dercle L <sup>[26]</sup> | Clin Cancer Res | 2020             | Retrospective; Multicenter | NSCLC          | 72          | 20               | Chest CT                          | Predict sensitivity to targeted-treatment                                 | Largest measurable lung lesion                     | Radiomics features*                                                                                                                | Random forest                                                   |
| Basler L <sup>[27]</sup> | Clin Cancer Res | 2020             | Retrospective              | Melanoma       | 112         | 112 <sup>#</sup> | <sup>18</sup> F-FDG PET/CT        | Distinguish hyper-progression in patients treated with immunotherapy      | All lesions                                        | Shape, intensity, and texture features                                                                                             | Logistic regression and regularized with L <sub>2</sub> penalty |
| Xu Y <sup>[28]</sup>     | Clin Cancer Res | 2019             | Retrospective; Multicenter | NSCLC          | 107         | 161              | Chest CT                          | Predict OS for patients with definitive RT and chemotherapy               | Manually defined the seed points of tumor as input | Deep learning features                                                                                                             | CNN and RNN                                                     |
| Song J <sup>[29]</sup>   | Clin Cancer Res | 2018             | Retrospective; Multicenter | NSCLC          | 1032        | 253              | Chest CE-CT                       | Discriminate patients with rapid and slow progression to EGFR-TKI therapy | Primary tumors                                     | Intensity, shape, and texture features                                                                                             | Cox regression model with LASSO penalty                         |
| Song J <sup>[30]</sup>   | JAMA Netw Open  | 2020             | Retrospective; Multicenter | NSCLC          | 145         | 320              | Chest CT with or without contrast | Predict survival for patients treated with TKI therapy                    | Whole CT images used as the input                  | 120-dimensional semantic features                                                                                                  | LASSO, Cox proportional hazards regression                      |
| Zhang N <sup>[31]</sup>  | Theranostics    | 2020             | Retrospective              | NSCLC          | 41          | 41               | <sup>18</sup> F-FDG PET/CT        | Predict treatment response to RT with concurrent chemotherapy             | Primary tumors; involved lymph nodes               | Morphology, boundary sharpness, intensity, and gray-level co-occurrence matrix features                                            | LASSO method                                                    |
| Jiang Y <sup>[32]</sup>  | Ebiomedicine    | 2018             | Retrospective; Multicenter | Gastric cancer | 228         | 1363             | Abdomen portal venous phase CE-CT | Predict survival when patients receive adjuvant chemotherapy              | Primary tumors                                     | Gray-level histogram, co-occurrence matrix, run-length matrix, absolute gradient, autoregressive model, wavelet transform features | LASSO method                                                    |

CE-CT, contrast enhanced-computerized tomography; FDG, fluorodeoxyglucose; PET, positron emission tomography; NSCLC, non-small cell lung cancer; EGFR, epidermal growth factor receptor; RT, radiation therapy; OS, overall survival; PD-L1, programmed death ligand-1; RAS, rat sarcoma viral oncogene homolog; BRAF, B-type Raf kinase; TKI, tyrosine kinase inhibitor; ResNet, residual neural network; ANN, artificial neural network; CNN, convolutional neural network; RNN, recurrent neural network; LASSO, least absolute shrinkage and selection operator. \* Radiomics features included: size-related, first order statistics, shape, surface shape, sigmoid functions, wavelets features, edge frequency features, fractal dimension, gray-tone difference matrix, Gabor energy, Laws' energy, laplacian of Gaussian, run-length features, spatial correlation, gray-level co-occurrence matrix feature classes. <sup>#</sup>nested cross-validation.

and the minimum sample size is 38 patients.<sup>[37]</sup> About 81.1% radiomics studies lack external validation.<sup>[37]</sup> However, Gilles *et al.* declared "in radiomics and elsewhere, the power of the predictive classifier model is dependent on having sufficient data".<sup>[7]</sup> In addition, a study demonstrated that even with cross-validation, studies with a small sample size are still unreliable.<sup>[38]</sup> Therefore, multi-center research is necessary to obtain sufficient sample size and avoid unstable and suboptimal results.

Some experts pointed out that there is an urgent need for standardized methods.<sup>[39]</sup> The standardized and unified methodology mainly refers to scanning predisposal, image acquisition, reconstruction, and *etc.* Different scanning protocols may introduce noise to images and influence the study results.<sup>[40]</sup> However, it is still challenging for radiology departments of different hospital/medical center to adopt and perform with the same scanning protocol nowadays. Meanwhile, some experts thought a diverse data set resembling the format and quality of data from the expected clinical setting is representative and ideal to train a model.<sup>[41]</sup> When a data set is large enough, noise caused by different scanning protocols could be reconciled.<sup>[41]</sup> As shown in Table 1, the definition and segmentation of lesions were different among studies. The segmentation process is subjective and time consuming, and required professional training, while neural network algorithm could address these challenges. Researchers proposed a novel 3D multi-attention guided multi-task learning network to realize automatic segmentation of gastric cancer.<sup>[42]</sup> Unlike supervised methods, unsupervised methods have produced means of segmenting lesions without labeling.<sup>[43]</sup> Repeatability assessment is required to test model robustness.<sup>[44]</sup> However, only 45.4% of radiomics studies conducted multiple segmentations by two readers, and only 0.1% of studies conducted segmentation at multiple time points.<sup>[37]</sup> The labor intensity of repeated manual segmentation is high, which may be the reason for insufficient repeatability evaluation. A study has attempted to test repeatability by both the second radiologist segmentation and the expansion of the original tumor mask with a morphological erosion operation algorithm on original segmentation to test the stability of the model.<sup>[17]</sup> This study provides, some references to reduce labor intensity in future studies. In addition, some specific issues still need to be solved, including interpretability of texture features, ethics and supervision of radiomics model.

It has been increasingly recognized that many clinical trials that involve AI interventions need to undergo rigorous evaluation to demonstrate their values in clinical outcome prediction. A study on reporting quality of radiomics on oncological studies showed that some researches were reported inadequately in terms of study objectives, blind assessment, sample size calculation, missing data, *etc.*,<sup>[37]</sup> which affects the reproducibility and clinical utility. The Consolidate Standard of Reporting Trails (CONSORT)-AI extension<sup>[45]</sup> published in 2020 has provided a standard study design and report guidance for studies with AI intervention in health care. It recommends 14 AI-related items (**Table 2**) that should be reported in addition to the routine items, which will promote the transparency and quality of the clinical trials involving AI techniques.

## SUMMARY

Technological innovation is endless. There have been substantial efforts in developing advanced radiological technology in the past decades. However, there still exist several limitations and hurdles that need to be overcome before widespread application in clinical practice. It is important that radiologists collaborate closely with both their clinical counterparts and AI researchers, and interact actively with interdisciplinary professionals. Through this way, the development of antineoplastic agents can be accelerated by transformation and application of radiomics in clinical practice, and ultimately benefit outcomes of cancer patients.

## Conflict of interests

None.

## REFERENCES

1. Smith BD, Kaufman MD, Lu WP, et al. Ripretinib (DCC-2618) is a switch control kinase inhibitor of a broad spectrum of oncogenic and drug-resistant KIT and PDGFRA variants. *Cancer Cell* 2019; 35(5):738-51.e9. doi:10.1016/j.ccell.2019.04.006.
2. Zhang X, Liang H, Li Z, et al. Perioperative or postoperative adjuvant oxaliplatin with S-1 versus adjuvant oxaliplatin with capecitabine in patients with locally advanced gastric or gastro-oesophageal junction adenocarcinoma undergoing D2 gastrectomy (RESOLVE): an open-label, superiority and non-inferiority, phase 3 randomised controlled trial. *Lancet Oncol* 2021;

**Table 2.** AI-related items that should be included in the report by the CONSORT-AI extension<sup>[45]</sup>

| Section                   | Item No. | CONSORT-AI item extensions                                                                                                                                                                |
|---------------------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title and Abstract        |          |                                                                                                                                                                                           |
| Title                     | 1a,b(i)  | Indicate that the intervention involves artificial intelligence/machine learning in the title and/or abstract and specify the type of model.                                              |
|                           | 1a,b(ii) | State the intended use of the AI intervention within the trial in the title and/ or abstract.                                                                                             |
| Introduction              |          |                                                                                                                                                                                           |
| Background and objectives | 2a(i)    | Explain the intended use for the AI intervention in the context of the clinical pathway, including its purpose and its intended users (e.g., healthcare professionals, patients, public). |
| Methods                   |          |                                                                                                                                                                                           |
| Participates              | 4a(i)    | State the inclusion and exclusion criteria at the level of participants.                                                                                                                  |
|                           | 4a(ii)   | State the inclusion and exclusion criteria at the level of the input data.                                                                                                                |
|                           | 4b       | Describe how the AI intervention was integrated into the trial setting, including any onsite or off-site requirements.                                                                    |
| Interventions             | 5(i)     | State which version of the AI algorithm was used.                                                                                                                                         |
|                           | 5(ii)    | Describe how the input data were acquired and selected for the AI intervention.                                                                                                           |
|                           | 5(iii)   | Describe how poor quality or unavailable input data were assessed and handled.                                                                                                            |
|                           | 5(iv)    | Specify whether there was human-AI interaction in the handling of the input data, and what level of expertise was required for users.                                                     |
|                           | 5(v)     | Specify the output of the AI intervention.                                                                                                                                                |
|                           | 5(vi)    | Explain how the AI intervention's outputs contributed to decision-making or other elements of clinical practice.                                                                          |
| Results                   |          |                                                                                                                                                                                           |
| Harms                     | 19       | Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, justify why not.                        |
| Other information         |          |                                                                                                                                                                                           |
| Funding                   | 25       | State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.                                                                |

CONSORT, Consolidate Standard of Reporting Trials; AI, artificial intelligence.

- 22(8):1081-92. doi:10.1016/S1470-2045(21)00297-7.
3. Gong J, Chehraz-Raffie A, Reddi S, et al. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. *J Immunother Cancer* 2018; 6(1):8. doi:10.1186/s40425-018-0316-z.
4. The Center for Drug Evaluation, National Medical Products Administration, China. The principles and procedures of clinical trial imaging endpoint process standards for antineoplastic agents. Available from: <https://www.cde.org.cn/main/att/download/ceb08a-90ca9af65d0bf1fa06235039e>. Released January 14, 2021; accessed August 15, 2021.
5. Choi H, Charnsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with ima-

- tinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007; 25(13):1753-9. doi:10.1200/JCO.2006.07.3049.
6. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* 2019; 69(2):127-57. doi:10.3322/caac.21552.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278(2):563-77. doi:10.1148/radiol.2015151169.
8. Erickson BJ, Korfiatis P, Akkus Z, et al. Machine learning for medical imaging. *Radiographics* 2017; 37(2):505-5. doi:10.1148/rg.2017160130.
9. Li Y, Cheng Z, Gevaert O, et al. A CT-based radiomics nomogram for prediction of human epidermal growth factor receptor 2 status in patients with gastric cancer. *Chin J Cancer Res* 2020; 32(1):62-71. doi:10.21147/

- j.issn.1000-9604.2020.01.08.
10. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019; 25(1):24-9. doi:10.1038/s41591-018-0316-z.
  11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521(7553):436-44. doi:10.1038/nature14539.
  12. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; 18(8):500-10. doi:10.1038/s41568-018-0016-5.
  13. Lee IS, Park YS, Lee JH, et al. Pathologic discordance of differentiation between endoscopic biopsy and postoperative specimen in mucosal gastric adenocarcinomas. *Ann Surg Oncol* 2013; 20(13):4231-7. doi:10.1245/s10434-013-3196-y.
  14. Xu F, Ma X, Wang Y, et al. CT texture analysis can be a potential tool to differentiate gastrointestinal stromal tumors without KIT exon 11 mutation. *Eur J Radiol* 2018; 107:90-7. doi:10.1016/j.ejrad.2018.07.025.
  15. 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):1180-91. doi:10.1016/S1470-2045(18)30413-3.
  16. Jiang Y, Wang H, Wu J, et al. Noninvasive imaging evaluation of tumor immune microenvironment to predict outcomes in gastric cancer. *Ann Oncol* 2020; 31(6):760-8. doi:10.1016/j.annonc.2020.03.295.
  17. Jiang Y, Liang X, Han Z, et al. Radiographical assessment of tumour stroma and treatment outcomes using deep learning: a retrospective, multicohort study. *Lancet Digit Health* 2021; 3(6):e371-e382. doi:10.1016/S2589-7500(21)00065-0.
  18. Mu W, Jiang L, Zhang J, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* 2020; 11(1):5228. doi:10.1038/s41467-020-19116-x.
  19. Rossi G, Barabino E, Fedeli A, et al. Radiomic detection of EGFR mutations in NSCLC. *Cancer Res* 2021; 81(3):724-31. doi:10.1158/0008-5472.CAN-20-0999.
  20. 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):2098-107. doi:10.7150/thno.48027.
  21. Shi R, Chen W, Yang B, et al. Prediction of KRAS, NRAS and BRAF status in colorectal cancer patients with liver metastasis using a deep artificial neural network based on radiomics and semantic features. *Am J Cancer Res* 2020; 10(12):4513-26.
  22. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* 2019; 30(6):998-1004. doi:10.1093/annonc/mdz108.
  23. Vaidya P, Bera K, Gupta A, et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in Stage I, II resectable Non-Small Cell Lung Cancer: a retrospective multi-cohort study for outcome prediction. *Lancet Digit Health* 2020; 2(3):e116-e128. doi:10.1016/S2589-7500(20)30002-9.
  24. 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):e001343. doi:10.1136/jitc-2020-001343.
  25. Dercle L, Lu L, Schwartz LH, et al. Radiomics response signature for identification of metastatic colorectal cancer sensitive to therapies targeting EGFR pathway. *J Natl Cancer Inst* 2020; 112(9):902-12. doi:10.1093/jnci/djaa017.
  26. 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):2151-62. doi:10.1158/1078-0432.CCR-19-2942.
  27. Basler L, Gabrys 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):4414-25. doi:10.1158/1078-0432.CCR-20-0020.
  28. Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 2019; 25(11):3266-75. doi:10.1158/1078-0432.CCR-18-2495.
  29. Song J, Shi J, Dong D, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. *Clin Cancer Res* 2018; 24(15):3583-92. doi:10.1158/1078-0432.CCR-17-2507.
  30. Song J, Wang L, Ng NN, et al. Development and validation of a machine learning model to explore tyrosine kinase inhibitor response in patients with stage IV EGFR variant-positive non-small cell lung cancer. *JAMA Netw Open* 2020; 3(12):e2030442. doi:10.1001/jamanetworkopen.2020.30442.
  31. Zhang N, Liang R, Gensheimer MF, et al. Early re-

- sponse evaluation using primary tumor and nodal imaging features to predict progression-free survival of locally advanced non-small cell lung cancer. *Theranostics* 2020; 10(25):11707-18. doi:10.7150/thno.50565.
32. Jiang Y, Chen C, Xie J, et al. Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. *EBioMedicine* 2018; 36:171-82. doi:10.1016/j.ebiom.2018.09.007.
33. Hudis CA. Trastuzumab--mechanism of action and use in clinical practice. *N Engl J Med* 2007; 357(1):39-51. doi:10.1056/NEJMr043186.
34. Li B, Jiang Y, Li G, et al. Natural killer cell and stroma abundance are independently prognostic and predict gastric cancer chemotherapy benefit. *JCI Insight* 2020; 5(9):e136570. doi:10.1172/jci.insight.136570.
35. Hegde PS, Karanikas V, Evers S. The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition. *Clin Cancer Res* 2016; 22(8):1865-74. doi:10.1158/1078-0432.CCR-15-1507.
36. Nishino M, Hatabu H, Hodi FS. Imaging of cancer immunotherapy: current approaches and future directions. *Radiology* 2019; 290(1):9-22. doi:10.1148/radiol.2018181349.
37. Park JE, Kim D, Kim HS, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol* 2020; 30(1):523-36. doi:10.1007/s00330-019-06360-z.
38. An C, Park YW, Ahn SS, et al. Radiomics machine learning study with a small sample size: Single random training-test set split may lead to unreliable results. *PLoS One* 2021; 16(8):e0256152. doi:10.1371/journal.pone.0256152.
39. Wang Y, Jin ZY. Radiomics approaches in gastric cancer: a frontier in clinical decision making. *Chin Med J (Engl)* 2019; 132(16):1983-9. doi:10.1097/CM9.0000000000000360.
40. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016; 6:23428. doi:10.1038/srep23428.
41. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019; 380(14):1347-58. doi:10.1056/NEJMr1814259.
42. Zhang Y, Li H, Du J, et al. 3D Multi-attention guided multi-task learning network for automatic gastric tumor segmentation and lymph node classification. *IEEE Trans Med Imaging* 2021; 40(6):1618-31. doi:10.1109/TMI.2021.3062902.
43. Aganj I, Harisinghani MG, Weissleder R, et al. Unsupervised medical image segmentation based on the local center of mass. *Sci Rep* 2018; 8(1):13012. doi:10.1038/s41598-018-31333-5.
44. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; 14(12):749-62. doi:10.1038/nrclinonc.2017.141.
45. Liu X, Cruz Rivera S, Moher D; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020; 26(9):1364-74. doi:10.1038/s41591-020-1034-x.