[¹⁸F]FDG-PET/CT radiomics and artificial intelligence in lung cancer
Technical aspects and potential clinical applications
Manafi-Farid, Reyhaneh; Askari, Emran; Shiri, Isaac; Pirich, Christian; Asadi, Mahboobeh; Khateri, Maziar; Zaidi, Habib; Beheshti, Mohsen

Published in:
Seminars in Nuclear Medicine

DOI:
10.1053/j.semnuclmed.2022.04.004

Publication date:
2022

Document version:
Final published version

Document license:
CC BY-NC-ND

Citation for published version (APA):

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying this open access version.

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk
Lung cancer is the second most common cancer and the leading cause of cancer-related death worldwide. Molecular imaging using \([^{18}F]\)fluorodeoxyglucose Positron Emission Tomography and/or Computed Tomography (\([^{18}F]\)FDG-PET/CT) plays an essential role in the diagnosis, evaluation of response to treatment, and prediction of outcomes. The images are evaluated using qualitative and conventional quantitative indices. However, there is far more information embedded in the images, which can be extracted by sophisticated algorithms. Recently, the concept of uncovering and analyzing the invisible data extracted from medical images, called radiomics, is gaining more attention. Currently, \([^{18}F]\)FDG-PET/CT radiomics is growingly evaluated in lung cancer to discover if it enhances the diagnostic performance or implication of \([^{18}F]\)FDG-PET/CT in the management of lung cancer. In this review, we provide a short overview of the technical aspects, as they are discussed in different articles of this special issue. We mainly focus on the diagnostic performance of the \([^{18}F]\)FDG-PET/CT–based radiomics and the role of artificial intelligence in non-small cell lung cancer, impacting the early detection, staging, prediction of tumor subtypes, biomarkers, and patient’s outcomes.

Introduction

Lung cancer is the second most common cancer and the leading cause of cancer-related death.\(^1\) Despite the decrease in mortality in recent decades,\(^2\) it remains a public health issue.\(^3\) The early diagnosis of the localized disease, accurate staging, response assessment, and prognostication are of paramount importance, which highly influence the treatment strategies and prognosis.\(^2,4,5\)

\([^{18}F]\)fluorodeoxyglucose positron emission tomography and/or computed tomography (\([^{18}F]\)FDG-PET/CT) is the widely accepted method for the non-invasive evaluation of non-small cell lung cancer (NSCLC) in different clinical settings.\(^3,8\) However, some limitations hinder \([^{18}F]\)FDG-PET/CT from becoming the ideal method. For example, inflammatory reactions in the thoracic region may cause false-positive findings, and small lesions and micro metastasis may be overlooked.\(^5\) Therefore,
there is a need for improvement of the current imaging techniques to fulfill the clinical requirements. 

[18F]FDG-PET/CT images are visually and semi-quantitatively evaluated for the clinical assessment. Nevertheless, it has been demonstrated that PET/CT images are not just photos. There are numerous other characteristics of the lesions and surrounding tissues, embedded in the images, which are imperceptible and are not calculated by the available practical software.10 These quantitative parameters, called features, can be extracted from the images and analyzed to predict different characteristics of the lesions. This process is known as radiomics, so-called the “more than meets the eye”11.

Radiomics uses digital data to define morphology, intensity, texture, etc. of the lesions and then correlates them with clinical, histological, and molecular findings. Over the last decade, [18F]FDG-PET/CT radiomics has been applied to differentiate malignant solitary pulmonary nodules (SPN) from benign lesions and increase staging accuracy, as well as to predict histology, tumor biomarkers, response to therapy, and prognosis. It should be noted that the number of the extracted features is too much to be assessed by the statistical analysis methods; therefore, artificial intelligence (AI) and machine learning (ML) are exploited to produce more accurate predictive models. There are several technical factors, such as harmonized data sets, which may limit the optimal implementation of radiomics in the clinical setting.12 To overcome these limitations, guidelines are developed for standardization,13 and also deep learning (DL) is increasingly employed to reduce some of the restrictions. However, there is still a gap between the studies and the translation of radiomics into clinical practice.

Several studies have investigated the value of radiomics in different aspects of NSCLC. Only 1% of the radiomics studies priorly were performed in the field of nuclear medicine.14 Encouragingly, this has recently reached 16%, considering that radiomics-like studies were present in nuclear medicine even before the word of “radiomics” itself came to existence.12 The main purpose of the current review is to provide an update on the recent status of [18F]FDG-PET/CT radiomics in the evaluation of NSCLC. First, we provide a brief review of the technical aspects. Afterwards, we summarize the recent studies assessing current applications of [18F]FDG-PET/CT radiomics and AI in NSCLC and seek an answer for “Do radiomics and AI increase the diagnostic performance or implication of [18F]FDG PET/CT in the era of precision medicine?”

Technical Aspect of Radiomics

Radiomics and Artificial Intelligence

Radiomics is a process to convert conventional images to minable data by extracting high dimensional quantitative semantic and/or agnostic features.15-18 Semantic features are defined as commonly used features for the region of interest (ROI) description by human observers.15-17 Agnostic features are those extracted by a computational process for ROI heterogeneity assessment. These features could be extracted using mathematical-based formula features (conventional radiomics features described by the image biomarker standardization initiative [IBSI]19) or deep convolutional neural network (CNN)-based features (extracted automatically through different convolutional layers).15-17 These features could be mine through the data mining process, employing different AI algorithms.20 AI tries to mimic human behaviors and ML, as a subset of AI, consists of different algorithms enabling computers to do this task without explicit programming.21, 22 ML algorithms are mostly applied to extracted features (agnostic and/or semantic) with feature selection or dimensionally reduction and regression or classification steps. DL is a ML algorithm that performs not only all these ML steps (feature selection, dimensionally reduction, regression, or classification) but also features extraction in one package.21

Radiomics’ Steps

Radiomics is a multidisciplinary (imaging technologist, medical physicist, radiologist, oncologist, statistician, computer scientist and data scientist) and multistep process, requiring different experts’ collaboration.15-17, 23, 24 The standard radiomics process includes data acquisition, image reconstruction, image segmentation, image pre-processing, feature extraction, feature selection and ML, as well as model evaluation,23 which we will discuss in following sections (Fig. 1).

Figure 1 Multiple steps of radiomics, including PET/CT image data acquisition, image reconstruction with applying multiple correction steps, volume of interests’ delineation, applying preprocessing and feature extraction, and machine learning steps (feature selection and classification/regression algorithm).
Data Acquisition
The image acquisition, the first step of radiomics study, is performed by technologists using different imaging modalities. Patients’ preparation and image acquisition should be performed in line with guidelines for PET/CT imaging to provide high-quality images, enabling reproducible and repeatable radiomics study. For PET image acquisition, the radiotracer activity should be administered based on weight or body mass index. The CT part of PET/CT scanning is usually performed for attenuation and/or scatter correction and anatomical localization and correlation. Using these data, PET images will be reconstructed by different mathematical algorithms to provide standard images for quantitative assessment. Different image artifacts could arise during PET/CT image acquisition. These artifacts could be due to PET image itself, including image noise due to radiotracer injection or failure in the scatter correction process, which results in halo artifact in regions with high activity. Some studies have addressed the image noise issue and employed advanced image processing using ML and DL to successfully decrease the image noise and increase the quality.

Other artifacts could be due to CT, such as truncation and metal artifact, which could change the quantitative information of PET data. Finally, the misregistration between PET and CT results in mislocation of lesions, for example, those close to the liver dome. The impact of motion should be considered for the radiomics analysis of lung tumors. Most recently, DL algorithms were proposed for direct attenuation and scatter corrections, bypassing the image reconstruction process with CT-based attenuation and scatter corrections, which potentially avoid or correct the mentioned artifacts. Image acquisition in PET (injected dose, time-of-flight [TOF], and time per bed position) and CT [kVp, mAs, and pitch] also impact on radiomics features. Proper values should be set to get high-quality images. In addition, their impact should be considered for providing repeatable and reproducible radiomics studies in multi-scanner and multi-centric settings.

Image Reconstruction
PET images could be reconstructed with different parameters, including reconstruction algorithm, TOF, point spread function, matrix size, iteration number, number of subsets and post-filtering. Previous studies have shown that reconstructions’ parameters could highly affect radiomics reproducibility in different imaging modalities. These effects are feature dependent, and each parameter has a different impact on radiomics features. Image reconstruction could be harmonized across different centers; however, different scanners from different providers have various sensitivity, which makes it infeasible to use the same set-up among imaging centers. Imaging data harmonization for different centers has been proposed to tackle this challenge, which is mentioned in the subsequent sections.

Image segmentation
Image segmentation is a crucial step for mathematical-based radiomic features extraction. Image segmentation could be performed manually (by nuclear medicine physicians, radiologists, oncologists), semi-automated (set initial seed for segmentation followed by manual editing), and fully automated by DL algorithms. Different segmentations performed on the same tumors result in different values of radiomics features, and the reproducibility of these features should be assessed precisely. Inter- and intra-observer variability and labor-intensive process of manual segmentation impede gathering large clean data sets for radiomic studies. DL-based algorithms have been recently used in PET image segmentation to automatically delineate ROIs, which outperform conventional PET image segmentation algorithms.

Image pre-processing
Image pre-processing for radiomics analysis is performed for various reasons, including noise reduction, computationally efficient feature extraction, and extracting different flavor of radiomics features. The image interpolation to isotropic voxel size should be done to extract rotationally invariant texture features, which could be performed by up or down sampling. Image discretization to fix bin number or fix bin width should be performed to normalize intensities; however, choosing the appropriate methods is an open question in radiomics studies. Recently, some authors investigated the impact of these parameters on the reproducibility of PET image radiomics features and showed that fixed bin width results in reproducible features in PET images. Various flavor of radiomics features could be extracted by using different filters, including wavelet (WL, applying either a high- or a low-pass filter in each of the three dimensions), Laplacian of Gaussian (LOG with different sigma values to extract fine, medium, and coarse texture), Exponential, Gradient, Logarithm, Square and Square Root scales for further radiomic investigation.

Feature extraction
Different types of features could be extracted from images, namely semantic and/or agnostic features. Semantic features are calculated by human observers describing ROI reporting different qualitative features that define the location, necrosis, spiculation and vascularity, etc. of tumors. Agnostic features are computed by mathematical-based formula description or deep CNN-based features. Mathematically-based radiomics features could be divided into shape-based, intensity and histogram-based, and texture features. Texture features consist of second-order, such as gray level co-occurrence matrix (GLCM), and high-order, including gray level run length matrix (GLRLM), gray level dependence matrix (GLDM), gray level size zone matrix (GLSZM), and neighboring gray tone difference matrix (NGTDM). As these features could be calculated employing different formula, using IBSI features is highly recommended to provide reproducible, and repeatable features. Different library and software such as Pyradiomics, SERA, LiFEx, CERR, MITK, QIFE, CaPTrk, RaCat, QuantImage, USZ, MIRP and QIFE have been evaluated for agreement with IBSI. Deep CNN-based features are automatically extracted using different convolutional layers. They could be applied to
images for feature extraction; however, there is no guideline for optimal number and order of layers and it highly depends on developer and task. The imbalance class in radiomics studies is another challenge, which could potentially bias the model and still provide high accuracy but low specificity or low sensitivity (depending on the class) that may not be useful in the clinical practice. Different approaches, such as data augmentation in image level or data sampling in features level have been proposed to address this issue.

Another subject is data sharing to build and evaluate a generalizable model due to legal and ethical problems. Most recently, federated learning algorithms have been proposed to build models without sharing data. These approaches have been developed for two different tasks in PET imaging and could potentially expand in radiomics studies. CT image information could be integrated into radiomics models in PET/CT studies using fusion in feature and fusion levels. Further studies could evaluate DL performance in PET/CT image fusion for radiomics models’ improvement.

Clinical Applications of [$^{18}$F]$^{18}$F-FDG-PET/CT Radiomics in Non-Small Cell Lung Cancer

[$^{18}$F]$^{18}$F-FDG-PET/CT is established as the standard imaging modality for the clinical management of NSCLC. However, there are still challenges for the interpretation of [$^{18}$F]$^{18}$F-FDG-PET/CT images, especially for differentiation of inflammatory from cancerous tissues and for treatment monitoring of the novel targeted therapies. Therefore, technical and quantitative assisting tools are needed for improving the diagnostic performance of [$^{18}$F]$^{18}$F-FDG-PET/CT in lung cancer for individualized disease management in different clinical scenarios, such as the early diagnosis, staging, prognostication, non-invasive evaluation of biomarkers and response assessment. To answer the clinical requirements, radiomics and AI are increasingly investigated in NSCLC in recent years. In the following sections, the value of [$^{18}$F]$^{18}$F-FDG-PET/CT radiomics is discussed in different clinical settings.

Pulmonary Nodules

The prevalence of SPN in normal populations ranges from 2%-24%, which increases to 17%-53% in patients with risk factors of occult malignancy. A non-negligible fraction of SPNs (ie, 1%-12%) harbors malignancy. Owing to more employment of CT, especially in the COVID-19 era, guidelines have been developed to prevent unnecessary measures for SPNs. A wide range of false-positive findings are detected on CT, leading to unavoidable harms and this is where radiomics comes to play.

There are multiple public SPN datasets to serve as a medium for radiomics surveys, of which the cancer imaging archive (TCIA) is one of the well-known repositories for [$^{18}$F]$^{18}$F-FDG-PET/CT images. Some studies have shown that radiomics using [$^{18}$F]$^{18}$F-FDG-PET/CT are superior to CT for the estimation of malignancy in SPNs (AUCs for

**Feature Selection and Machine Learning Models**

Large numbers of features could be extracted from ROIs; however, all these features are not informative for a specific task, and ML models (which perform classification, regression, or time to event task) would be highly prone to overfitting. Dimensionality reduction and feature selection could be performed in supervised (filter-based, model-based and hybrid), semi-supervised and unsupervised approaches. Different ML algorithms, including rule-based model, linear and nonlinear regression, neural networks, support vector, naïve bays and ensemble learning-based have been developed. These algorithms could be applied to features for classification (binary and multiclass classification), regression (ie, age prediction) and time to event prediction (survival analysis). There is no “one fits all ML model” for a specific task, ML models parameters and hyperparameters should be tuned based on task, and different ML models could be evaluated to reach the optimal ones.

**ML Model Evaluation**

Radiomics models could be evaluated using different metrics (depending on the task) through different approaches. These approaches include data splitting to train and/or validation and/or test sets, one-leave-out, cross validation, bootstrapping, one-leave-center-out (in case of multicentric study). For the regression task, different metrics, including mean error (ME), mean absolute error (MAE), relative error (RE%), absolute relative error (ARE%), and normalized version of these metrics could be calculated. For time to event tasks, c-index and hazard ratio are the model parameter evaluation metrics. For classification tasks, different metrics should be reported to assess models’ performance, namely, accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC), positive predictive value (PPV), and negative predicted value (NPV). For classification task different metrics should be reported to assess models performances, specifically in case of unbalance classes, NPV and PPV should be reported to assess power model in rare cases as we can get high accuracy and AUC but missing rare cases. Using external validation set is highly recommended for model generalizability assessments.

**Challenges and New Horizons**

Different steps of radiomics such as image acquisition, reconstruction and segmentation could highly affect radiomics features value, resulting in non-repeatable and non-reproducible features. Different studies have evaluated the impact of these parameters on radiomics features’ repeatability and reproducibility. Recently, harmonization approaches have been proposed in the feature domain (ie, ComBat) and image-level (ie, generative adversarial network) to tackle variability due to image acquisition, scanner, and reconstruction setting.
conventional radiomics: 0.809-0.940 for \({}^{[18]}\text{F}\)FDG-PET/CT vs 0.646-0.908 for CT; and AUCs for DL: 0.877 for \({}^{[18]}\text{F}\)FDG-PET/CT vs 0.817 for CT).\(^{93-96}\) Interestingly, a recent multicentric prospective head-to-head comparison of these two modalities showed even more striking results. The AUCs for dynamic contrast-enhanced CT and \({}^{[18]}\text{F}\)FDG-PET/CT were 0.62 and 0.80, respectively, yet complementary to each other (combined AUC = 0.90).\(^{97}\)

Given the limited spatial resolution of \({}^{[18]}\text{F}\)FDG-PET/CT, small pulmonary nodules may not be detectable on \({}^{[18]}\text{F}\)FDG-PET and therefore, \({}^{[18]}\text{F}\)FDG-PET/CT is usually reserved for SPNs > 8 mm.\(^{86}\) However, it seems that \({}^{[18]}\text{F}\)FDG-PET-derived radiomics features can be more sensitive (94% vs 58%) and accurate (93% vs 76%) than visual analysis for the detection of small lesions.\(^{98}\) Moreover, emerging data suggest that ML can be helpful to reduce the noise, scan time, injected activity, and reconstruction time without a meaningful drop in the AUCs.\(^{73, 99}\) Also, DL approaches using newer image reconstruction methods may impact the diagnosis of small SPNs.\(^{100}\)

The inherent nature of \({}^{[18]}\text{F}\)FDG-PET/CT leads to well-known false-positive findings, interfering with the diagnosis of lung cancer.\(^{101}\) This challenge has been addressed by radiomics studies trying to reduce the false-positive rate. Radiomics can more accurately differentiate benign inflammatory lesions, such as tuberculosis (AUC = 0.889-0.93),\(^{102, 104}\) pneumonia (accuracy = 82.5%),\(^{103}\) and radiation pneumonitis (accuracy = 85%)\(^{100}\) from lung cancer.\(^{84, 107}\) Also, some studies using texture analysis have suggested a possible added benefit of dual time-point imaging to further enhance the AUCs of radiomics (from 0.52-0.75 to 0.63-0.87).\(^{108, 109}\) A nomogram has also been developed using \({}^{[18]}\text{F}\)FDG-PET/CT radiomics combined with manual diagnosis, which has decreased the false-positive rate of manual diagnosis by 21.5% (AUC = 0.92).\(^{104}\) Moreover, CNN method has shown a 93% reduction in false-positive results with a cost of 7% reduction in the sensitivity.\(^{110}\) On the other hand, some studies have shown that radiomics-based nomograms using conventional radiomics did not outperform nuclear medicine experts in the validation cohort, while DL algorithms marginally surpassed physicians.\(^{103, 111}\)

In a study to differentiate tuberculosis from lung adenocarcinoma, Hu et al. showed that the radiomics model outperformed the clinical model with marginal inferiority to the combined model (AUCs of 0.889, 0.644 and 0.909, respectively).\(^{102}\) Another study demonstrated insignificantly lower, yet complementary, performance of \({}^{[18]}\text{F}\)FDG-PET-CT compared to CT-based features for differentiation of tuberculosis from lung cancer (AUC: 0.91 vs 0.85, \(P = 0.1554\)).\(^{103}\)

The summary of the studies is provided in Table 1. Although there are potential benefits for the prediction of malignancy in SPNs, there are some drawbacks.\(^{112}\) and AI is still immature in this field to be clinically applied. In fact, a comprehensive review revealed that most radiomics studies in this domain lack relevant comparator or independent and/or external validation of the models.\(^{113}\) Also, for ML, 3D approaches still suffer from sufficient sample size for such purposes.\(^{114, 116}\) Noteworthy, along with the ability of automatic detection, DL algorithms seem to have higher discriminative power, possibly due to bypassing the segmentation and considering the tumor surrounding and whole-body data. Future studies should focus on the algorithms of DL combined with clinical information to increase the diagnostic performance of AI in the detection of nodules and differentiation of malignant from benign lesions.

**Histologic Subtype Differentiation**

Given the invasive nature of tissue biopsy, inadequacy and/or non-feasibility of sampling in some cases, and also the unmet need to differentiate the pathologic subtypes for treatment planning, radiomics come to play for the non-invasive differentiation of the histologic tumor subtypes.\(^{117}\)

In this regard, different studies have shown a discriminative role for radiomics and ML models.\(^{118, 119}\) Table 2 provides the summary of the investigations. Early radiomics studies pointed out that texture features had a role in distinguishing adenocarcinoma (ADC) from Sqamous cell carcinoma (SqCC).\(^{119}\) Some authors showed that \({}^{[18]}\text{F}\)FDG-PET radiomics features and ML models outperformed those of CT (AUC: 0.80-0.83 vs 0.69-0.79).\(^{120, 121}\) Also, DL methods have been employed, showing superiority over conventional radiomics for differentiation of ADC from SqCC (AUC: 0.841 vs 0.794).\(^{117}\)

Additionally, some authors implemented clinical factors in the prediction models, demonstrating the additive value of combined models over radiomics-only models (AUCs of 0.78-0.98 vs 0.70-0.94).\(^{121-123}\) Ren et al. showed that the model based on both clinical factors and tumor markers is superior to imaging-based models for discrimination of subtypes; however, the combined model had a higher AUC of 0.90.\(^{121}\)

There are also some efforts to differentiate primary and metastatic lung lesions using radiomics (AUC: 0.61-0.97).\(^{120, 124}\) Again, \({}^{[18]}\text{F}\)FDG-PET-based radiomics were superior to CT-based features (AUC: 0.57 vs 0.88).\(^{120}\) A hybrid approach using ML models could optimally classify primary and metastatic lung lesions using \({}^{[18]}\text{F}\)FDG-PET (AUC = 0.983) and CT (AUC = 0.828).\(^{125}\) Another study showed that fused signature combining \({}^{[18]}\text{F}\)FDG-PET/CT and clinical findings achieved the highest diagnostic accuracy (AUC = 0.953).\(^{126}\) Moreover, the ComBat harmonization method has been applied to multicenter data, improving the predictive performance of PET and fused PET/CT models.\(^{82}\)

Overall, the \({}^{[18]}\text{F}\)FDG-PET/CT-based radiomics seem promising for the differentiation of tumor subtypes, especially using DL and implementing clinical factors in the prediction models. The non-invasive evaluation is of particular importance in patients with large or unresectable lesions or those with known other malignancies. Apparently, due to higher prevalence, most studies have focused on ADC and SqCC. Differentiating other subtypes is also worth investigating in future studies using multiclass classification ML algorithms.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Pt No.</th>
<th>Aim</th>
<th>Reference standard</th>
<th>Segmentation method</th>
<th>Classifier</th>
<th>Dataset classification*</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen107</td>
<td>2017</td>
<td>85</td>
<td>Malignant vs Benign</td>
<td>Pathology or follow-up</td>
<td>Manual</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.91</td>
</tr>
<tr>
<td>Guo106</td>
<td>2017</td>
<td>40</td>
<td>Discrimination of lung cancer from pneumonia</td>
<td>-</td>
<td>Manual</td>
<td>ML</td>
<td>-</td>
<td>Accuracy = 85%</td>
</tr>
<tr>
<td>Suga238</td>
<td>2021</td>
<td>63</td>
<td>Discrimination of lung cancer from pneumonia</td>
<td>Pathology</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>One dataset - type 1a</td>
<td>AUC = 0.82-0.83</td>
</tr>
<tr>
<td>Watanabe105</td>
<td>2018</td>
<td>20</td>
<td>Discrimination of lung cancer from pneumonia</td>
<td>Pathology or follow-up</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>-</td>
<td>Accuracy = 82.5%</td>
</tr>
<tr>
<td>Wu115</td>
<td>2018</td>
<td>2,789,675</td>
<td>Malignant vs Benign</td>
<td>-</td>
<td>-</td>
<td>ML</td>
<td>-</td>
<td>Accuracy = 77%</td>
</tr>
<tr>
<td>Chen108</td>
<td>2019</td>
<td>116</td>
<td>Malignant vs Benign</td>
<td>Pathology or follow-up</td>
<td>Manual</td>
<td>Statistical analysis</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.99</td>
</tr>
<tr>
<td>Kang104</td>
<td>2019</td>
<td>268</td>
<td>Reducing PET/CT false positive rate</td>
<td>Pathology or follow-up</td>
<td>Manual and Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.98</td>
</tr>
<tr>
<td>Nakajo109</td>
<td>2019</td>
<td>59</td>
<td>Malignant vs Benign</td>
<td>Pathology</td>
<td>Manual and automatic</td>
<td>Statistical analysis</td>
<td>One dataset - type 1a</td>
<td>AUC = 0.98</td>
</tr>
<tr>
<td>Zhang94</td>
<td>2019</td>
<td>135</td>
<td>Malignant vs Benign</td>
<td>Pathology or follow-up</td>
<td>Manual</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.887</td>
</tr>
<tr>
<td>Hu102</td>
<td>2020</td>
<td>235</td>
<td>Malignant vs Benign</td>
<td>Pathology</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.889</td>
</tr>
<tr>
<td>Chen64</td>
<td>2021</td>
<td>317</td>
<td>Malignant vs Benign</td>
<td>Pathology</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 1b</td>
<td>AUC = 0.727</td>
</tr>
<tr>
<td>Du103</td>
<td>2021</td>
<td>174</td>
<td>Malignant vs Benign</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.93</td>
</tr>
<tr>
<td>Park95</td>
<td>2021</td>
<td>359</td>
<td>Malignant vs Benign</td>
<td>Pathology</td>
<td>-</td>
<td>DL</td>
<td>-</td>
<td>AUC = 0.837 and 0.877</td>
</tr>
<tr>
<td>Shao114</td>
<td>2021</td>
<td>106</td>
<td>Malignant vs Benign</td>
<td>Pathology or follow-up</td>
<td>Semiautomatic</td>
<td>DL</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.97</td>
</tr>
<tr>
<td>Zhou125</td>
<td>2021</td>
<td>769</td>
<td>Discrimination of lung cancer from metastasis</td>
<td>Pathology</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 1b/2a</td>
<td>AUC = 0.983</td>
</tr>
<tr>
<td>Zhang111</td>
<td>2022</td>
<td>174</td>
<td>Malignant vs Benign</td>
<td>Pathology</td>
<td>Manual</td>
<td>DL/ ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.84</td>
</tr>
</tbody>
</table>

AUC, Area under curve; DL, Deep learning; ML, Machine learning; PET/CT, Positron emission tomography/computed tomography; Pt No., Number of patients. All studies were retrospective.
*Based on TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) classification.
Table 2  The Summary of Studies Differentiating the Histologic Subtypes of Non-Small Cell Lung Cancer Using Radiomics and Artificial Intelligence

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Pt No.</th>
<th>Aim</th>
<th>Reference standard</th>
<th>Segmentation method</th>
<th>Classifier</th>
<th>Dataset Classification</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma</td>
<td>2018</td>
<td>341</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.89</td>
</tr>
<tr>
<td>Kirienko</td>
<td>2018</td>
<td>534</td>
<td>Differentiation between primary and metastatic lung lesions</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.91</td>
</tr>
<tr>
<td>Li</td>
<td>2018</td>
<td>207</td>
<td>Classification of pulmonary nodules/ Predicting histological subtypes</td>
<td>Pathology</td>
<td>-</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC ≥ 0.907</td>
</tr>
<tr>
<td>Sha</td>
<td>2019</td>
<td>100</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Temporal external validation - type 2b</td>
<td>AUC = 0.781</td>
</tr>
<tr>
<td>Han</td>
<td>2021</td>
<td>1419</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>ML, DL</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.903</td>
</tr>
<tr>
<td>Koyasu</td>
<td>2020</td>
<td>188</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>-</td>
<td>AUC = 0.843</td>
</tr>
<tr>
<td>Yan</td>
<td>2020</td>
<td>445</td>
<td>Differentiation of subtypes and primary and metastatic lung lesions</td>
<td>Pathology</td>
<td>Automatic</td>
<td>ML</td>
<td>Random split-sample 1b/2a</td>
<td>AUC = 0.98 and 0.99</td>
</tr>
<tr>
<td>Zhou</td>
<td>2021</td>
<td>769</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>ML</td>
<td>Random split-sample - type 1b/2a</td>
<td>AUC = 0.897</td>
</tr>
<tr>
<td>Ren</td>
<td>2021</td>
<td>315</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.90</td>
</tr>
<tr>
<td>Ji</td>
<td>2021</td>
<td>253</td>
<td>Differentiation of subtypes</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>ML</td>
<td>Temporal external validation - Type 2b</td>
<td>AUC = 0.978-0.989</td>
</tr>
</tbody>
</table>

AUC, Area under curve; DL, Deep learning; ML, Machine learning; Pt No., Number of patients.
All studies are retrospective.
*Based on TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) classification.

Table 3  Summary of Studies Evaluating the Staging in Non-Small Cell Lung Cancer Patients Using Radiomics and Artificial Intelligence

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Pt. No.</th>
<th>Aim</th>
<th>Reference standard</th>
<th>Segmentation method</th>
<th>Classifier</th>
<th>Dataset Classification</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gao</td>
<td>2015</td>
<td>132</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.689</td>
</tr>
<tr>
<td>Coroller</td>
<td>2016</td>
<td>108</td>
<td>Prediction of DM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AUC = 0.64</td>
</tr>
<tr>
<td>Wu</td>
<td>2016</td>
<td>101</td>
<td>Prediction of DM</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>ML</td>
<td>Temporal external validation - Type 2b</td>
<td>AUC = 0.80</td>
</tr>
<tr>
<td>Kirienko</td>
<td>2017</td>
<td>31</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>Statistical analysis</td>
<td>One dataset - type 1a</td>
<td>For some features: p&lt;0.05</td>
</tr>
<tr>
<td>Wang</td>
<td>2017</td>
<td>168</td>
<td>LN-staging</td>
<td>Imaging and pathology</td>
<td>Manual</td>
<td>ML, DL</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.91</td>
</tr>
<tr>
<td>Kirienko</td>
<td>2018</td>
<td>472</td>
<td>T-staging</td>
<td>Imaging and pathology</td>
<td>Semi-automatic</td>
<td>DL</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.68</td>
</tr>
<tr>
<td>Lyu</td>
<td>2020</td>
<td>130</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>-</td>
<td>ML</td>
<td>-</td>
<td>AUC = 0.917</td>
</tr>
<tr>
<td>Tau</td>
<td>2020</td>
<td>264</td>
<td>Prediction of LN and DM and follow-up</td>
<td>Imaging, pathology, and manual</td>
<td>Semi-automatic</td>
<td>DL</td>
<td>Resampling - type 1b</td>
<td>LN: AUC = 0.8</td>
</tr>
<tr>
<td>Chang</td>
<td>2021</td>
<td>528</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.94</td>
</tr>
<tr>
<td>Taralli</td>
<td>2021</td>
<td>540</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>DL</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.789</td>
</tr>
<tr>
<td>Wallis</td>
<td>2021</td>
<td>125</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>DL</td>
<td>Different scanner external validation - type 2b</td>
<td>Sensitivity = 0.88</td>
</tr>
<tr>
<td>Yoo</td>
<td>2021</td>
<td>980</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Semi-automatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.85</td>
</tr>
<tr>
<td>Zheng</td>
<td>2021</td>
<td>716</td>
<td>LN-staging</td>
<td>Pathology</td>
<td>Manual</td>
<td>ML</td>
<td>Split-sample - type 2a</td>
<td>AUC = 0.80</td>
</tr>
</tbody>
</table>

AUC, Area Under Curve; DL, Deep learning; DM, Distant metastasis; LN, Lymph node; ML, Machine learning; Pt No., Number of patients.
All studies were retrospective.
*Based on TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) classification.
Staging

T-staging

T-staging is usually performed using CT, and can be improved by [18F]FDG-PET data (Table 3). A few studies in this subtopic exists for the role of [18F]FDG-PET/CT radiomics. For example, Kirienko et al. designed a CNN-based algorithm and classified NSCLC as T1/2 or T3/4 with an AUC of 0.68. The uptake value (SUV) of [18F]FDG-PET are used to predict LN staging to the models to discover the additive value of metabolic parameters for N-staging (AUCs: combined = 0.838-0.872 vs 0.82-0.872 vs CT = 0.822-0.828). Although the models showed an acceptable predictive power, it was marginally higher than the accuracy of conventional parameters. The corresponding values for diffusion-weighted magnetic resonance imaging imaging were 72% and 97%, respectively. Therefore, invasive pathology logic evaluation is used for mediastinal N-staging with its inherent disadvantages. The discrimination of N0 and N1 disease is especially important when the standard surgery is not planned and the LNs are not dissected. Therefore, the additive value of radiomics is being evaluated in mediastinal N-staging.

N-staging

Lymph node (LN) staging is a crucial step for optimum treatment in NSCLC and is still a challenging issue. The current imaging modalities do not provide enough accuracy for this purpose, especially in endemic areas of granulomatous diseases. The patient-based and node-based sensitivities of [18F]FDG-PET/CT are approximately 62%-67% with the specificities of about 87%-93%. The corresponding values for diffusion-weighted magnetic resonance imaging imaging are 72% and 97%, respectively. Therefore, invasive pathology logic evaluation is used for mediastinal N-staging with its inherent disadvantages. The discrimination of N0 and N1 disease is especially important when the standard surgery is not planned and the LNs are not dissected. Therefore, the additive value of radiomics is being evaluated in mediastinal N-staging.

Traditionally, the Hounsfield unit of CT and standardized uptake value (SUV) of [18F]FDG-PET are used to predict LN metastasis. The use of ML-based classification in this subtopic roots back to before applying radiomics, when simple metabolic parameters, such as SUVmax, were implemented in non-radiomics studies to build predictive ML models (AUCs: 0.886-0.962). Moreover, some studies evaluated the CT-based radiomics and added only the SUVmax to the models to discover the additive value of metabolic parameters for N-staging (AUCs: combined = 0.838-0.872 vs CT = 0.822-0.828). Although the models showed an acceptable predictive power, it was marginally higher than the accuracy of conventional parameters. The corresponding values were 0.73 for short-axis diameter ≥ 1.0 cm and 0.82 for SUVmax ≥ 2.5.

Using [18F]FDG-PET/CT radiomics, the reported accuracies for N-staging are rather wide. In an early study, ML was not superior to traditional SUVmax (0.652 vs 0.579-0.689, respectively). However, in another study, the [18F]FDG-PET/CT-based model resulted in an AUC of 0.80 compared to 0.61 for physicians. Also, Wang et al. claimed that the accuracy of CNN for N-staging is not significantly higher than that of assessed by experts (AUC = 0.91 and accuracy = 0.86 vs accuracy = 0.82). It seems that AI models show higher sensitivity while physicians’ reports have higher specificity with comparable accuracy for predicting histologic LN status. One of the interesting results of Yoo et al.’s study was that the sensitivity of ML was approximately twice higher than that of physicians in LNs with SUVmax < 3.5, but the specificity was still lower. Some authors have incorporated the clinical information into [18F]FDG-PET/CT radiomics, reporting higher performance for combined models. Yoo et al. reported an AUC of 0.85 for the combined model vs 0.75 for physicians.

Patients with more hypermetabolic tumors have a higher chance of occult LN metastasis. The radiomics of the primary tumor has also been evaluated to predict occult LN involvement, reaching an AUC of 0.78 with ML and 0.80 with CNN. Of interest, a model combining [18F]FDG-PET with breath-hold thin-slice chest CT achieved the AUC of 0.917 in stage I adenocarcinoma.

One of the main challenges is the robustness of the features or models using different acquisition systems or reconstruction methods. Wallis et al. used CNN to predict LN metastasis in the mediastinum. They applied transfer learning to the second set of data obtained from another acquisition system, which increased the sensitivity from 0.53-0.88. However, the false-positive rate was also increased (from 0.24-0.69).

In summary (Table 3), there are limitations for the assessment of N-staging using [18F]FDG-PET/CT radiomics, such as the presence of reactive LNs or difficulty in histologic evaluation of each LN. The diagnostic performance of AI in the prediction of LN involvement seems only slightly higher than that of experts. In this regard, a recent systematic review studied the performance of the radiomics models for staging purposes. The AUCs, sensitivities and specificities of the studies were 0.64-0.94, 52%-99%, and 60%-99%, respectively. The authors concluded that heterogeneous study designs and lack of prospective validation in most studies preclude using radiomics for current clinical decisions. Additionally, there are limitations for most studies, excluding small or inactive lesions, although the small size and low metabolism of the metastasis are the main challenges for N-staging. Future studies should focus on the role of automatic detection, DL, and combined models-incorporating all LN features, primary tumor characteristics and clinical information.

M-staging

Current imaging modalities can determine the M-stage with high accuracy. Even though [18F]FDG-PET/CT is the modality of choice in this realm, with the exception of central nervous system metastasis, any suspicious foci of metastatic involvement should be ascertained by using additional imaging or biopsy. Radiomics has been implemented to also predict distant metastasis. In early radiomics studies, the features of the primary tumor barely predicted distant metastasis (AUC = 0.64-0.71). Interestingly, even CNN did not improve the performance for prediction for distant metastasis at diagnosis (AUC = 0.71). Whether [18F]FDG-PET/CT radiomics may one day obviate the need for complementary imaging or biopsy is yet to be determined.

Prediction of Biomarkers

Lung cancer is a heterogeneous tumor, and various responsible gene mutations have been detected so far. Tumor genotype is translated into its phenotype. Evaluation
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Pt No.</th>
<th>Aim</th>
<th>Reference standard</th>
<th>Segmentation method</th>
<th>Classifier</th>
<th>Dataset classification</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yip</td>
<td>2017</td>
<td>348</td>
<td>Predicting EGFR and KRAS mutations</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>One dataset - type 1a</td>
<td>AUC = 0.67 and 0.54, retrospectively</td>
</tr>
<tr>
<td>Zhang</td>
<td>2018</td>
<td>180</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Manual</td>
<td>ML</td>
<td>Split-sample</td>
<td>AUC = 0.8725</td>
</tr>
<tr>
<td>Novikov</td>
<td>2019</td>
<td>84</td>
<td>Prediction of tumor grade</td>
<td>Pathology</td>
<td>Semiautomatic</td>
<td>DL</td>
<td>Random split-sample - type 2a</td>
<td>Accuracy = 91-100%; r = 0.592, p = 0.028</td>
</tr>
<tr>
<td>Moon</td>
<td>2019</td>
<td>176</td>
<td>Prediction of mutation burden</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>One dataset - type 1a</td>
<td></td>
</tr>
<tr>
<td>Li</td>
<td>2019</td>
<td>115</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Manual and Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.805</td>
</tr>
<tr>
<td>Moitra</td>
<td>2019</td>
<td>211</td>
<td>Automated grading</td>
<td>Histology</td>
<td>-</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.96</td>
</tr>
<tr>
<td>Jiang</td>
<td>2019</td>
<td>80</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.953</td>
</tr>
<tr>
<td>Mu</td>
<td>2020</td>
<td>616</td>
<td>Treatment guidance (TKI vs ICI)</td>
<td>Genomic analysis, follow-up and imaging</td>
<td>Semiautomatic</td>
<td>DL</td>
<td>External validation -type 3</td>
<td>AUC = 0.81</td>
</tr>
<tr>
<td>Zhang</td>
<td>2020</td>
<td>248</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.87</td>
</tr>
<tr>
<td>Nair</td>
<td>2020</td>
<td>50</td>
<td>Predicting EGFR mutation/ Differentiating mutation exons</td>
<td>Genomic analysis</td>
<td>Manual</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>AUC = 0.87 and 0.86, respectively</td>
</tr>
<tr>
<td>Shiri</td>
<td>2020</td>
<td>150</td>
<td>Predicting EGFR and KRAS mutations</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.82 and 0.80, respectively</td>
</tr>
<tr>
<td>Shiri</td>
<td>2022</td>
<td>136</td>
<td>Predicting EGFR and KRAS mutations</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.94 and 0.93, respectively</td>
</tr>
<tr>
<td>Sanduleanu</td>
<td>2020</td>
<td>221</td>
<td>Prediction of tumor hypoxia</td>
<td>Hypoxia-PET</td>
<td>Manual</td>
<td>ML</td>
<td>External validation -type 3</td>
<td>AUC = 0.73</td>
</tr>
<tr>
<td>Whi</td>
<td>2020</td>
<td>64</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Manual</td>
<td>Statistical analysis</td>
<td>One dataset - type 1a</td>
<td>OR = 4.06-4.57</td>
</tr>
<tr>
<td>Yang</td>
<td>2020</td>
<td>174</td>
<td>Predicting EGFR mutation/ Differentiating mutation exons</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.71 and 0.73, respectively</td>
</tr>
<tr>
<td>Zhang</td>
<td>2020</td>
<td>173</td>
<td>Predicting EGFR mutation/ Differentiating mutation exons</td>
<td>Genomic analysis</td>
<td>Manual and automatic</td>
<td>ML</td>
<td>Random split-sample - type 1b</td>
<td>AUC = 0.827 and 0.661, respectively</td>
</tr>
<tr>
<td>Liu</td>
<td>2021</td>
<td>148</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.87</td>
</tr>
<tr>
<td>Chang</td>
<td>2021</td>
<td>583</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Manual and Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 1b/2a</td>
<td>AUC = 0.84</td>
</tr>
<tr>
<td>Yin</td>
<td>2021</td>
<td>301</td>
<td>Predicting EGFR mutation</td>
<td>Genomic analysis</td>
<td>Manual</td>
<td>DL</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.84</td>
</tr>
<tr>
<td>Li</td>
<td>2021</td>
<td>255</td>
<td>Predicting PD-L1 expression</td>
<td>Genomic analysis</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample – type 2a/1b</td>
<td>AUC = 0.84 and 0.762, respectively</td>
</tr>
<tr>
<td>Nie</td>
<td>2021</td>
<td>272</td>
<td>Predicting lymphovascular invasion</td>
<td>Histology</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>External validation -type 3</td>
<td>AUC = 0.838</td>
</tr>
<tr>
<td>Chang</td>
<td>2021</td>
<td>526</td>
<td>Predicting ALK rearrangement status</td>
<td>IHC</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>AUC = 0.88</td>
</tr>
</tbody>
</table>

ALK, Anaplastic lymphoma kinase; AUC, Area Under Curve; DL, Deep learning; EGFR, Epidermal growth factor receptor; ICI, Immune checkpoint inhibitor; IHC, Immunohistochemistry; KRAS, Kirsten rat sarcoma virus; ML, Machine learning; OR, Odds Ratio; PD-L1, Programmed death-ligand 1; PET, Positron emission tomography; Pt No., Number of patients; TKI, Tyrosine kinase inhibitor.

All studies were retrospective.

*Based on TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) classification.
of mutations is becoming an inevitable step in the management of NSCLC.\textsuperscript{161} impacting the treatment options. Additionally, some features, such as tumor grade,\textsuperscript{162} lymphovascular invasion (LVI),\textsuperscript{4} or hypoxia,\textsuperscript{163} influence the outcome. Therefore, predicting the presence of such biomarkers using the standard diagnostic imaging would be substantially advantageous, which is being investigated using $^{[18]}$F-FDG-PET/CT radiomics (Table 4).

A study failed to demonstrate a significant relation between radiomics features and the mutation burden.\textsuperscript{164} The other showed modest relations between a number of mutations and $^{[18]}$F-FDG-PET/CT radiomics.\textsuperscript{165} To begin with, mutations in epidermal growth factor receptor (EGFR) have a significant role in the development of NSCLC.\textsuperscript{166} Those with EGFR mutation respond better to tyrosine kinase inhibitors (TKIs).\textsuperscript{167} Therefore, the non-invasive prediction of EGFR mutation is tempting. Tumors with EGFR mutation tend to have more heterogeneity,\textsuperscript{168, 169} which might be captured by radiomics. Employing $^{[18]}$F-FDG-PET/CT conventional radiomics and ML for the prediction of EGFR mutation, a wide range of predictive power has been reported (AUC = 0.5-0.87), mostly around 0.75-0.80.\textsuperscript{168, 170-172} $^{[18]}$F-FDG-PET/CT-based features seemed superior over CT and PET-only radiomics (AUC = 0.80-0.84, 0.67-0.72, and 0.74-0.79, respectively).\textsuperscript{168, 173} Also, ML models outperformed conventional factors (AUCs = 0.82 vs 0.75).\textsuperscript{171}

To further improve the accuracy, a number of scholars added clinical data in the predictive models, showing that the combined models (AUC = 0.82-0.87) surpassed imaging-based models (AUC = 0.68-0.77)\textsuperscript{168, 169, 176} and the clinical model alone (AUC = 0.69).\textsuperscript{169} However, the accuracies of DL and DL plus clinical data were similar in another survey (AUC = 0.81 vs 0.84).\textsuperscript{174} Moreover, in a study, $^{[18]}$F-FDG-PET/CT data combined with visual features of the images yielded a high AUC of 0.953.\textsuperscript{175} In an attempt to increase performance, Shiri et al. applied feature harmonization, which proved to be feature-dependent, and slightly increased AUC for the prediction of EGFR mutation in different image modality including PET, CT and fused PET/CT image using ML algorithm (AUC increased from 0.87-0.90 to 0.92-0.94).\textsuperscript{71}

There are two major EGFR mutations, 19DEL and 21L858R, which influence outcome and treatment options.\textsuperscript{179} In this regard, $^{[18]}$F-FDG-PET/CT radiomics may discriminate 19DEL from 21L858R (AUC of 0.73-0.87).\textsuperscript{170, 180} Liu et al. predicted the presence of 19DEL and 21L858R with AUC of 0.77 and 0.92, respectively.\textsuperscript{181} On the other hand, Zhang et al. incorporated clinical information in the predictive model and reported that only one $^{[18]}$F-FDG-PET-based feature could differentiate 19DEL from 21L858R mutation with low predictive potential (AUC = 0.66).\textsuperscript{182}

Kristen rat sarcoma viral (KRAS) also plays a significant role in NSCLC.\textsuperscript{166} The predictive role of radiomics for KRAS mutation status is controversial. None of the $^{[18]}$F-FDG-PET-based features predicted KRAS mutation in a study by Yip et al.\textsuperscript{172} However, Shiri et al. demonstrated that a few features are predictive for KRAS mutation (AUC = 0.71).\textsuperscript{171} Also, they reported that the best model for predicting KRAS mutation is the CT-based ML model.\textsuperscript{171} In a most recent study by Shiri et al. AUCs of 0.91–0.94 were reported for KRAS mutation classification using harmonized PET/CT fused image.\textsuperscript{71}

The level of programmed death-ligand 1 (PD-L1) expression is an important factor for choosing immunotherapy in NSCLC.\textsuperscript{183} The potential of radiomics to predict PD-L1 expression status is evaluated in a few studies. The PD-L1 expression levels >1% and >50% were predicted with AUCs of 0.762 and 0.814, respectively.\textsuperscript{184} Similarly, Zhou et al. assessed tumor microenvironment immune types (ie, a combination of PD-L1 and CD8+ expression) and reached similar results with AUCs of 0.794, 0.699 and 0.811 for $^{[18]}$F-FDG-PET/CT-radiomics, clinical and combined models, respectively.\textsuperscript{185} However, Jiang et al., showed that CT-derived models (AUC 0.80–0.97) are superior to $^{[18]}$F-FDG-PET-based models (0.61–0.75) for the prediction of PD-L1 expression.\textsuperscript{186}

Anaplastic lymphoma kinase (ALK) inhibitors are also used for the treatment of NSCLC patients with ALK rearrangement or positive for ROS1 (c-ros oncogene 1) or RET (rearranged during transfection) fusion.\textsuperscript{187, 188} Change et al. employed radiomics to predict ALK mutation status and reported that $^{[18]}$F-FDG-PET/CT radiomics predicted ALK mutation with an AUC of 0.86.\textsuperscript{187} The result was comparable with the model based on combined $^{[18]}$F-FDG-PET/CT plus clinical data (AUC = 0.88).\textsuperscript{187} The predictive features were more selected from CT features, possibly due to the higher resolution of CT compared to PET (1 mm vs 5 mm).\textsuperscript{187} Moreover, Yoon et al. distinguished ALK, ROS1, or RET fusion-positive NSCLC from fusion-negative tumors with sensitivity and specificity of 73% and 70%, respectively.\textsuperscript{188}

Higher tumor grade\textsuperscript{192} and LVI\textsuperscript{14} are associated with a poorer prognosis. Novikov et al. employed radiomics-based models using different segmentation methods to predict tumor grade.\textsuperscript{162} None of the individual features were predictive.\textsuperscript{162} However, incorporating all features in predictive models, all three tumor grades were discriminated with overall accuracies of 71 to 100%.\textsuperscript{162} Also, Moitra et al. reported that fuzzy-rough nearest neighbor classifier, among other ML methods, provides higher performance for grading.\textsuperscript{189} Moreover, Nie and colleagues could predict LVI using imaging-based features, with the CT-based radiomics plus SUV$_{max}$ providing the highest AUC (0.79).\textsuperscript{143}

The proliferation rate is another prognostic factor in NSCLC.\textsuperscript{190} In this regard, Palumbo et al. evaluated SUV$_{max}$ and diameter of lesions on $^{[18]}$F-FDG-PET/CT images and applied ML to predict Ki-67 index, showing that the combination of SUV$_{max}$ and lesion diameter could predict Ki-67 of > 25% with an accuracy of 82%.\textsuperscript{190}

Hypoxia is another biomarker which causes resistance to treatment in NSCLC. Sanduleanu et al. predicted hypoxic fraction > 20% using PET and CT-derived radiomics with AUC of 0.71-0.82 in both lung and head and neck cancers.\textsuperscript{163}

All in all, there is a plethora of factors and biomarkers impacting NSCLC patients’ outcomes. Radiomics seems to predict EGFR mutation with acceptable but not ideal
The evaluation of biomarkers. Future studies can shed light on the role of radiomics in the non-invasive evaluation of different biomarkers in NSCLC.

Prognosis and Outcome

Predicting the outcome is an essential step for the patients' management. Various prognostic factors, such as stage, biomarkers, etc., have been identified. Also, the conventional semi-quantitative metabolic parameters on [18F]FDG-PET/CT have shown prognostic value. It is crucial to detect patients with a higher probability of recurrence who may benefit from more aggressive local or systemic treatments. The role of radiomics have also been investigated in this scenario (Table 5).

Early studies pointed out the inverse relation between tumor heterogeneity and overall survival (OS). For example, in a study, the reduced heterogeneity was associated with a response to erlotinib. Also, texture features such as entropy, dissimilarity, coarseness, contrast, and busyness were predictors of outcome in patients undergoing radiation therapy (RT) or chemoradiotherapy. Others found Asphericity to be a predictor of outcome. These features had additional prognostic value to metabolic tumor volumes (MTVs) and/or clinical variables. Later, radiomics-based models were successfully developed to predict more outcome measures, namely OS, disease-specific survival and regional control.

Attempts to predict OS or progression-free survival (PFS) using textural features showed promising results (AUCs = 0.665-0.762). The majority of studies showed an additive prognostic value of radiomics to clinical risk factors. Even in some studies, radiomics features outperformed the combined clinical and radiomics data (AUCs of 0.68-0.75 vs 0.61-0.65). Also, combined PET/CT models seem superior to each modality alone. In contrast, some studies have reported that radiomics features were incapable of predicting PFS in patients undergoing curative RT, and some have questioned the independent prognostic value of [18F]FDG-PET only images.

Surgery

Surgery is the treatment of choice in stage I/II NSCLC, and some patients with stage IIIA disease or even some cases with oligometastasis. However, 30%-55% of patients experience local/distant recurrence after curative surgery. The value of [18F]FDG-PET/CT textural features for better stratification of the patients undergoing surgery has been addressed in a number of studies.

Ahn et al. assessed the value of [18F]FDG-PET/CT radiomics for the prediction of disease-free survival (DFS) in patients undergoing surgery. They predicted recurrence with AUCs of 0.871-0.956 using ML. However, the clinical stage was a more powerful factor for the prediction of DFS. Kirienko et al. showed that the AUC of the CT-based model was higher than that of [18F]FDG-PET (0.75 vs 0.68). Unexpectedly, the AUCs dropped after consideration of clinical findings (0.61 vs 0.65). Using two different scanners and the lack of some important clinical data limited their study. On the other hand, Christie et al. added clinical stage to [18F]FDG-PET/CT radiomics model and achieved an AUC of 0.79.

“Extracting more information from medical images” needs one to think creatively and outside of the box. Radiomics studies usually focus on the target lesion while capturing more data from the penumbra and/or peritumoral area or even outside the tumor boundaries may provide additional information of prognostic significance. For example, Mattonen et al. evaluated the peritumoral region radiomics to predict recurrence and/or progression. The clinical stage was the best predictor of recurrence and/or progression (AUC = 0.68). However, adding MTV to peritumoral data empowered the predictive model (AUC = 0.74). Also, further addition of bone marrow uptake radiomics improved performance (AUC = 0.78). Others have tried to predict cachexia as an independent prognostic factor. This kind of approach to radiomics highlights the need for an interdisciplinary approach, an essential requirement for radiomics studies.

Radiation Therapy

RT is another treatment option for NSCLC patients. Unfortunately, no universal model beyond staging exists for prognostication of patients undergoing RT. Radiomics and pattern recognition using textural features may have a role to address the need for prediction of local recurrence, nodal failure or distant metastasis following RT.

In an early study, tumor- and LN-conventional features were used to predict recurrence after RT; however, the study failed to show a significant predictive power for local recurrence, tumor and LN features predicted overall recurrence (AUC = 0.69). Li et al. predicted OS and nodal failure (AUCs = 0.64 and 0.66, respectively). Furthermore, Okinomou et al. showed the superiority of [18F]FDG-PET/CT radiomics over simple SUVmax measurements for the prediction of OS and recurrence. The performance was also improved involving clinical information (AUC = 0.97), as well as genetic and follow-up data (AUC = 0.79) during therapy (known as delta features).

For distant metastasis prediction, radiomics studies showed the additive role of radiomics to the simple imaging or clinical parameters. Also, the ML model based on both [18F]FDG-PET and CT achieved higher AUC compared to...
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Pt No.</th>
<th>Aim (prediction of) Treatment</th>
<th>Segmentation method</th>
<th>Classifier</th>
<th>Data classification</th>
<th>Follow-up time (months)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tixier201</td>
<td>2014</td>
<td>102</td>
<td>OS, recurrence-free survival</td>
<td>Surgery, RT, chemotherapy, combination therapy</td>
<td>Manual and automatic</td>
<td>Statistical analysis</td>
<td>36.6 (Median)</td>
<td>Some features predicted survival with p &lt; 0.05</td>
</tr>
<tr>
<td>Nakajo211</td>
<td>2018</td>
<td>55</td>
<td>OS</td>
<td>Surgery</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>23 (Median)</td>
<td>Multivariate analysis: Stage HR = 1.62, IV HR = 6.19</td>
</tr>
<tr>
<td>Harmon212</td>
<td>2019</td>
<td>64</td>
<td>OS, PFS</td>
<td>Surgery</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mattonen227</td>
<td>2019</td>
<td>227</td>
<td>DFS</td>
<td>Surgery</td>
<td>Manual and semiautomatic</td>
<td>ML</td>
<td>Temporal split-sample - type 2b</td>
<td>41 (Median)</td>
</tr>
<tr>
<td>Mattonen221</td>
<td>2019</td>
<td>291</td>
<td>Time to recurrence/ progression</td>
<td>Surgery</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>External validation - type 3†</td>
<td>32-50 (Median varies per dataset)</td>
</tr>
<tr>
<td>Christie222</td>
<td>2021</td>
<td>135</td>
<td>Recurrence</td>
<td>Surgery</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Split-sample - type 2</td>
<td>-</td>
</tr>
<tr>
<td>Pyka224</td>
<td>2015</td>
<td>45</td>
<td>LR and DSS</td>
<td>SBRT</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Statistical analysis</td>
<td>21.4 (Median)</td>
</tr>
<tr>
<td>Lovinfosse227</td>
<td>2016</td>
<td>63</td>
<td>OS, DFS, DSS</td>
<td>SBRT</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>27.1 (Median)</td>
<td>HR for dissimilarity = DDS: 0.822 and DFS: 0.834</td>
</tr>
<tr>
<td>Hao223</td>
<td>2017</td>
<td>48</td>
<td>Distant failure</td>
<td>SBRT</td>
<td>-</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>18 (Median)</td>
</tr>
<tr>
<td>Takeda231</td>
<td>2017</td>
<td>26</td>
<td>LC, OS, PFS</td>
<td>SBRT</td>
<td>-</td>
<td>ML</td>
<td>Resampling - type 1a</td>
<td>36 (Median)</td>
</tr>
<tr>
<td>Zhou232</td>
<td>2018</td>
<td>52</td>
<td>Distant failure</td>
<td>SBRT</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>18 (Median)</td>
</tr>
<tr>
<td>Li232</td>
<td>2018</td>
<td>100</td>
<td>OS, Nodal failure</td>
<td>SBRT</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 2</td>
<td>18 (Median)</td>
</tr>
<tr>
<td>Oikonomou211</td>
<td>2018</td>
<td>150</td>
<td>RC, disease control, RFF, DSS, and OS</td>
<td>SBRT</td>
<td>Manual</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>27 (Median)</td>
</tr>
<tr>
<td>Dissaux233</td>
<td>2020</td>
<td>87</td>
<td>LC</td>
<td>SBRT</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>External validation - type 3</td>
<td>21.1-25.5 (Median varies per dataset)</td>
</tr>
<tr>
<td>Cook235</td>
<td>2013</td>
<td>53</td>
<td>treatment response, OS, PFS, local PFS</td>
<td>Chemoradiotherapy</td>
<td>Manual</td>
<td>Statistical analysis</td>
<td>21.2 (Median)</td>
<td>AUCs for prediction of response: contrast = 0.82</td>
</tr>
<tr>
<td>Apostolova207</td>
<td>2014</td>
<td>60</td>
<td>OS, PFS</td>
<td>Surgery ± (chemo) radiotherapy or chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>10.1 (Median)</td>
<td>PFS: asphericity (HR = 3.60) and solidity (HR = 2.11)</td>
</tr>
<tr>
<td>Fried208</td>
<td>2015</td>
<td>195</td>
<td>OS</td>
<td>Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>&gt; 1 years; for living patients = 37 (Median)</td>
</tr>
<tr>
<td>Li231</td>
<td>2015</td>
<td>25</td>
<td>Local and overall relapse</td>
<td>Chemoradiotherapy</td>
<td>Manual</td>
<td>ML</td>
<td>Resampling - type 1b†</td>
<td>26 (Median)</td>
</tr>
<tr>
<td>Ohri208</td>
<td>2016</td>
<td>201</td>
<td>OS</td>
<td>Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>24</td>
</tr>
<tr>
<td>Dong212</td>
<td>2016</td>
<td>58</td>
<td>OS, PFS</td>
<td>Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>60 (Median)</td>
<td>respectively, PFS: HR = 0.476, OS: HR = 0.519</td>
</tr>
<tr>
<td>Kirienko214</td>
<td>2017</td>
<td>295</td>
<td>DFS</td>
<td>RT or chemotherapy</td>
<td>Semiautomatic</td>
<td>Statistical analysis</td>
<td>20.1-20.5 (for CT or PET dataset, Median)</td>
<td>AUC = 0.68</td>
</tr>
<tr>
<td>Luo215</td>
<td>2018</td>
<td>118</td>
<td>LC and Radiation Pneumonitis</td>
<td>(Chemoradiotherapy</td>
<td>-</td>
<td>ML</td>
<td>Temporal split-sample - type 2b</td>
<td>61-65 (Median varies per group)</td>
</tr>
<tr>
<td>Jensen210</td>
<td>2018</td>
<td>79</td>
<td>OS</td>
<td>Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>One dataset - type 1a</td>
<td>22 (Median)</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Pt No.</td>
<td>Aim (prediction of) Treatment</td>
<td>Segmentation method</td>
<td>Classifier</td>
<td>Data classification*</td>
<td>Follow-up time (months)</td>
<td>Results</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>--------</td>
<td>-------------------------------</td>
<td>---------------------</td>
<td>------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Arshad</td>
<td>2019</td>
<td>358</td>
<td>OS (Chemo)radiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Random and nonrandom split-sample - type 2</td>
<td>22 (Median)</td>
<td>HR = 1.61</td>
</tr>
<tr>
<td>Krarup</td>
<td>2019</td>
<td>233</td>
<td>PFS Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>One dataset - type 1a</td>
<td>631.5 days (Mean)</td>
<td>No radiomics features predicted PFS</td>
</tr>
<tr>
<td>van Timmeren</td>
<td>2019</td>
<td>138</td>
<td>OS Chemoradiotherapy</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>3.4-6.9 years (Median varies per dataset)</td>
<td>AUC ranged from 0.66 to 0.89</td>
</tr>
<tr>
<td>Ahn</td>
<td>2019</td>
<td>93</td>
<td>DFS Surgery ± chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>45 (Median)</td>
<td>AUC for RF = 0.956</td>
</tr>
<tr>
<td>Astaraki</td>
<td>2019</td>
<td>30</td>
<td>OS Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>2 years</td>
<td>AUC &gt; 0.90</td>
</tr>
<tr>
<td>Konert</td>
<td>2020</td>
<td>362</td>
<td>OS Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>External validation - type 3</td>
<td>17-24 (Median varies per dataset)</td>
<td>AUC &gt; 0.51 to 0.59 less than clinical model</td>
</tr>
<tr>
<td>Zhang</td>
<td>2020</td>
<td>82</td>
<td>OS Chemoradiotherapy</td>
<td>Manual</td>
<td>ML</td>
<td>Random split-sample - type 2a</td>
<td>1.9-2 years (Median varies per group)</td>
<td>AUC &gt; 0.77-0.79</td>
</tr>
<tr>
<td>Carles</td>
<td>2021</td>
<td>48</td>
<td>OS, LR, DM Chemoradiotherapy</td>
<td>Manual and Semiautomatic</td>
<td>ML</td>
<td>External validation - type 3†</td>
<td>-</td>
<td>AUC &gt; 0.63</td>
</tr>
<tr>
<td>Moran</td>
<td>2021</td>
<td>39</td>
<td>OS Chemoradiotherapy</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AUC &gt; 0.82-0.83</td>
</tr>
<tr>
<td>Park</td>
<td>2018</td>
<td>182</td>
<td>PFS Chemoradiotherapy</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Temporal external validation - type 3</td>
<td>-</td>
<td>AUC = 0.662</td>
</tr>
<tr>
<td>Mu</td>
<td>2020</td>
<td>194</td>
<td>DCB, PFS, OS ICI</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>Resampling - type 1b</td>
<td>21.4 (Median)</td>
<td>AUC &gt; 0.90</td>
</tr>
<tr>
<td>Park</td>
<td>2020</td>
<td>181</td>
<td>CAS (marker for prediction of OS, PFS, treatment response) ICI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AUCs = 0.74-0.88</td>
</tr>
<tr>
<td>Valentinuzzi</td>
<td>2020</td>
<td>30</td>
<td>Treatment response (OS &gt; median of 14.9 months) ICI</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>AUC &gt; 0.90</td>
</tr>
<tr>
<td>Polverari</td>
<td>2020</td>
<td>57</td>
<td>Treatment response ICI</td>
<td>Manual and semiautomatic</td>
<td>ML</td>
<td>One dataset - type 1a</td>
<td>10 (Median)</td>
<td>Association between several features with progressive disease</td>
</tr>
<tr>
<td>Mu</td>
<td>2021</td>
<td>697</td>
<td>DCB, PFS, OS ICI</td>
<td>Semiautomatic</td>
<td>DL</td>
<td>External validation - type 3</td>
<td>-</td>
<td>AUC &gt; 0.82</td>
</tr>
<tr>
<td>Mu</td>
<td>2021</td>
<td>210</td>
<td>Cachexia ICI</td>
<td>Semiautomatic</td>
<td>ML</td>
<td>External validation - type 3</td>
<td>26 (Median)</td>
<td>AUCs &gt; 0.74</td>
</tr>
<tr>
<td>Shao</td>
<td>2021</td>
<td>250</td>
<td>PFS</td>
<td>Manual</td>
<td>ML</td>
<td>External validation - type 3</td>
<td>&gt; 2 years</td>
<td>AUC &gt; 0.60-0.71</td>
</tr>
</tbody>
</table>

AUC: Area under curve; CXR, Chest X-ray; CAS, Cytolytic activity score; DCB, Durable clinical benefit; DFS, Disease-free survival; DL, Deep learning; DM, Distant metastasis; DSS, Disease-specific survival; HILAE, High-intensity large-area emphasis; HR, Hazard ratio; ICI, Immune checkpoint inhibitor; LC, Local control; LR, Local recurrence; ML, Machine learning; OS, Overall survival; PFS, Progression-free survival; PET/CT, Positron emission tomography/computed tomography; Pt No., Number of patients; RC, Regional control; RFP, Recurrence-free probability; RT, Radiotherapy; SBRT, Stereotactic body radiation therapy; TKI, Tyrosine kinase inhibitor.

*Based on TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) classification.

†At least one cohort included prospectively gathered cases.
to either alone. Moreover, DL outperformed ML models for predicting distant disease. 

Differentiating radiation pneumonitis from residual/recurrent disease to avoid delay in treatment is a challenging issue after RT. These may have a similar appearance on CT. Also, increased metabolism due to inflammation is common in [18F]FDG-PET/CT images, which may decrease the accuracy of SUVs. To distinguish these two entities, showed that two conventional radiomics features outperformed SUV (AUC = 0.83-0.82 vs 0.63-0.64) but not MTV (AUC = 0.86). 

ChemoRadiotherapy 
Concurrent chemoradiotherapy is usually recommended for stage II LN-positive or stage III diseases. It may make patients prone to toxicity from intensive therapy without having a significant survival benefit. For patients receiving chemoradiation, several factors are of prognostic significance. Yet, current clinical and imaging parameters are still imperfect in terms of prognostic power. The added value of radiomics signature to SUVs or clinical factors in NSCLC patients undergoing chemoradiation has been described (AUCs: 0.72-0.79). A study could not build a successful radiomics-based model. However, a multicenter study evaluated the pre-treatment [18F]FDG-PET/CT radiomics features for the prediction of survival in stage I-II NSCLC. The radiomics features-derived hazard ratio (HR) was 1.61 (CI 95%; 1.16-2.24), while SUV-related parameters were not predictive. Another smaller study emphasized the added value of combined [18F]FDG-PET radiomics with conventional prognostic factors to conventional prognostic factors-alone or CT radiomics for the prediction of OS (AUCs of 0.82, 0.68, and 0.62, respectively). Similarly, focused on a new partitioning method to predict OS. Their ML model outperformed conventional radiomics features (AUC = 0.90 vs. 0.71). Using the changes in radiomics features (delta features) in the follow-up studies after chemoradiation, patients with increasing homogeneity in the primary tumor had a higher rate of local recurrence. 

Immunotherapy 
Immune checkpoint inhibitors (ICIs) have an emerging role as consolidation therapy after chemoradiation in stage II/III NSCLC. They can also be employed with/without chemotherapy for the first-line treatment of metastatic patients. To predict response to ICIs, a study found a good AUC for [18F]FDG-PET/CT radiomics (0.81) and showed its additive value to the nomogram (AUCs of 0.77 and 0.80 for the prediction of OS and PFS, respectively). Furthermore, Mu et al. used DL and predicted PD-L1 expression status (AUC ≥0.82), durable clinical benefit (AUC = 0.87), PFS (AUC = 0.77), and OS (AUC = 0.70). 

Following immunotherapy, pseudoprogression is likely, and imaging criteria may lag months to define the real progression warranting additional imaging. Radiomics may help predict the tumor response. Given that tumor heterogeneity can predict disease progression in these patients, a study compared immunotherapy radiomics (so-called “iRADiomics”) with iRECIST (immunotherapy response evaluation criteria in solid tumors), showing a higher AUC for the prediction of response to treatment with pembrolizumab (0.90 vs 0.79-0.86).

For the prediction of response to neoadjuvant immunotherapy in early-stage NSCLC, [18F]FDG-PET-derived radiomics features were uneventful in a study with small numbers of patients. However, employing tumor immune microenvironment, cytolytic activity score was predictive using [18F]FDG-PET/CT-based DL (AUCs = 0.74-0.88), which also correlated with PFS and OS. Additionally, responders to ICI were discriminated from non-responders (P = 0.005). Additionally, some authors tried to predict cachexia (as a factor for resistance to immunotherapy, AUC ≥0.74) or severe immune-related adverse events (AUC = 0.88).

Discussion and Conclusion 
An exponentially increasing number of studies are being published evaluating the role of AI in medical imaging. The preliminary results show that [18F]FDG-PET/CT-based radiomics provides additional valuable information in NSCLC patients. To recapitulate, AI provides the possibility of automatic detection and also denoising and increasing the quality of studies with low-dose imaging, theoretically opening a horizon for screening of malignancies. In addition, radiomics may differentiate malignant pulmonary nodules and the sub-types of the primary lesions, especially when combined with clinical data, slightly better than physicians, moving one more step toward the non-invasive characterization of the lesions. The superiority of radiomics for the T- and M-staging has not been proved in the limited number of surveys.
Likewise, the prediction of LN involvement seems only slightly higher than that of experts. Additionally, the expanding field of biomarkers in oncology has been addressed. Most studies have shown promising potential for the prediction of EGFR mutation status, which may influence the decision for targeted therapies in the future. However, other biomarkers have scarcely been assessed. Moreover, it is crucial to predict the response to therapy in the early stages of treatment to timely adjust the management. In this regard, a large proportion of studies have reported the promising role of radiomics in predicting outcome as the endpoint of cancer management of studies have reported the promising role of radiomics in predicting outcome as the endpoint of cancer management.

Nevertheless, the immediate physiological relevance of radiomics is not yet defined. A recent systematic review in this regard also concluded that the quality score of the studies is low, reflecting a lack of reproducibility. There are a few studies evaluating the prognostic value of [18F]FDG-PET/CT radiomics to predict outcomes after surgery or chemoradiation, showing a wide range of accuracies. There are known prognostic clinical factors. They should also be implemented in the predictive models to augment the predictive performance. Moreover, incorporating data from the peritumoral regions or even outside the tumor boundaries may provide additional prognostic information. Given the whole-body evaluation with DL, it may show higher performance in this regard. Also, there is a debate regarding the role of radiomics in prognosis after RT. Two recent systematic reviews addressed the utility of radiomics in this regard, showing a modest predictive value for conventional radiomics for OS. However, DL seems more promising, which should be employed in future studies. These emerging pieces of evidence are expressed as a “significant potential” of radiomics in the recent EANM/SNMMI/ESTRO guidelines on the role of [18F]FDG-PET/CT in RT planning. Moreover, targeted therapies will have prominent roles in the management of NSCLC in the future. Considering the costs and adverse events of targeted therapy, it is better to define which patients will have a durable clinical benefit. Incorporating radiomics may enhance the prognostic power of [18F]FDG-PET/CT and may impact the decision making.

A comprehensive review on radiomics biomarkers in the field of immunotherapy is available, yet data on [18F]FDG-PET radiomics is not solid and warrants further improvement. Finally, delta features seem to possess prognostic significance in response evaluation. However, there are challenges in maintaining the consistency of parameters in different studies and robustness of the relevant features.

In conclusion, [18F]FDG-PET/CT radiomics seems to be advantageous for the evaluation of NSCLC in different settings. However, all fields suffer from shared drawbacks, i.e. standardized imaging, the method of radiomics implementation, and reporting, limit comparative studies and translation of radiomics into clinical practice. The repeatability and reproducibility of radiomics features should be assessed for robust radiomics modeling. Harmonization, data augmentation and federated learning algorithm could be employed to tackle acquisition/reconstruction variability, imbalance classes and data sharing challenges in the clinic, respectively. Efforts have been made for standardization and improving the quality of reports, including the development of IBSI, TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis), and RQS (radiomics quality score), yet to be implemented in all published studies. Moreover, DL algorithms are increasingly employed and seemingly have equal or superior performance compared to conventional statistical analysis and ML models. Moreover, adding the clinical information into predictive models, mimicking the human decision approach, may improve the efficacy.

The wide implementation of AI in medical imaging research is to reach the unfulfilled dream of developing a rapid, one-step, diagnostic “machine.” It seems that there is still a long way to reach that goal. Also, whether virtual biopsy one day will become a reality remains to be answered. However, by publishing standard reports, experts can use AI to improve their reports and enhance patients’ management. To accelerate the clinical use of radiomics, future studies should comply with the guidelines. Introducing the potential applications, advantages and limitations of AI in scientific meetings, seminars, and webinars would help increase awareness.

### Funding

This research received no external funding.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Acknowledgments

We wish to express our deep appreciation and gratitude to Milad Enferadi, Research Center for Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, for helping in literature search and providing the manuscripts.

### References

13. The image biomarker standardisation initiative: Creative Commons
10. Mana
20. Youse
19. Zwanenburg A, Valli
15. Gillies RJ, Kinahan PE, Hricak H: Radiomics: Images are more than
23. van Timmeren JE, Cester D, Tanadini-Lang S, et al: Radiomics in medi-
25. Vaništekníková J, Črnko L, Dooms C, et al. 2nd ESMO Consensus con-
29. Xie X, Zhao H, Guo Y, et al: A PET/CT nomogram incorporating SUV-
24. Eberhardt WE, De Ruyscher D, Weder W, et al: 2nd ESMO Consen-
23. 31:126-133, 2021
22. 11:91, 2020
21. 58:365-366, 2017
29. 58:365-366, 2017
27. 16:595-608, 2017
26. 27:370-2744, 2021
25. 16:449-469, 2021
24. 101770, 2020
23. 104665, 2021
22. 17:183-212, 2022
20. 15:93, 2015
19. 312:170-173, 2018
18. 63:122-133, 2004
17. 33:156-161, 2005. quiz 163-4
16. 57:1692-1698, 2016
15. 14:e0221877, 2019
14. 19:184-186, 2019
13. 41:917, 2020
12. 27:449-4507, 2019
11. 5:102, 2018
10. 15:93, 2015
9. 46:872-883, 2021
8. 136:104665, 2021
7. 28:2730-2744, 2021
6. 50:917, 2020
5. 30:12, 2020
4. 4:0221877, 2019
3. 46:872-883, 2021
2. 25:1462-1474, 2014
1. 38:102, 2018
0. 312:122-133, 2004
malignancy in solid pulmonary nodules based on a population with high prevalence of malignancy. Clin Lung Cancer 21:47-55, 2020


162. Novikov M: Multiparametric quantitative and texture 18F-FDG PET/CT analysis for primary malignant tumour grade differentiation. Eur Radiol Exp 3:48, 2019


