Application of PET/CT-based deep learning radiomics in head and neck cancer prognosis: a systematic review

Shuyan Li\textsuperscript{a,b}, Jinghua Liu\textsuperscript{c,d}, Zhongxiao Wang\textsuperscript{e}, Zhendong Cao\textsuperscript{f}, Yifan Yang\textsuperscript{g}, Bingzhen Wang\textsuperscript{h}, Shiqi Xu\textsuperscript{i}, Lijun Lu\textsuperscript{j}, M. Iqbal Saripan\textsuperscript{k}, Xiaolei Zhang\textsuperscript{l}, Xianling Dong\textsuperscript{a,b,*, Dong Wen\textsuperscript{i, *}}

\textsuperscript{a}Hebei International Research Center for Medical-Engineering, Chengde Medical University, Hebei, China
\textsuperscript{b}Department of Biomedical Engineering and Hebei Key Laboratory of Nerve Injury and Repair, Chengde Medical University, Hebei, China
\textsuperscript{c}Department of Nursing, Chengde Central Hospital, Hebei, China
\textsuperscript{d}Department of Nursing, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia
\textsuperscript{e}Department of Radiology, The Affiliated Hospital of Chengde Medical University, Hebei, China
\textsuperscript{f}Faculty of Environment and Life, Beijing University of Technology, Beijing, China
\textsuperscript{g}School of Biomedical Engineering and Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China
\textsuperscript{h}Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia
\textsuperscript{i}Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing, China

\*Correspondence: wencangdong@163.com (D. Wen); dongxl_cdmc@163.com (X. Dong)

Received: 14 July 2022; Revised: 26 September 2022; Accepted: 12 October 2022

Published online: 23 November 2022

DOI 10.15212/RADSCI-2022-0006

Abstract

Background: Radiomics and deep learning have been widely investigated in the quantitative analysis of medical images. Deep learning radiomics (DLR), combining the strengths of both methods, is increasingly used in head and neck cancer (HNC). This systematic review was aimed at evaluating existing studies and assessing the potential application of DLR in HNC prognosis.

Materials and methods: The PubMed, Embase, Scopus, Web of Science, and Cochrane databases were searched for articles published in the past 10 years with the keywords “radiomics,” “deep learning,” and “head and neck cancer” (and synonyms). Two independent reviewers searched, screened, and reviewed the English literature. The methodological quality of each article was evaluated with the Radiomics Quality Score (RQS). Data from the studies were extracted and collected in tables. A systematic review of radiomics prognostic prediction models for HNC incorporating deep learning techniques is presented.

Result: A total of eight studies, published in 2012–2022, with a varying number of patients (59–707 cases), were included. Each study used deep learning; three studies performed automatic segmentation of regions of interest (ROI), and the Dice score range for automatic segmentation was 0.75–0.81. Four studies involved extraction of deep learning features, one study combined different modality features, and two studies performed predictive model building. The range of the area under the curve (AUC) was 0.84–0.96, the range of the concordance index (C-index) was 0.72–0.82, and the range of model accuracy (ACC) was 0.72–0.96. The median total RQS for these studies was 13 (10–15), corresponding to a percentage of 36.11% (27.78%–41.67). Low scores were due to a lack of prospective design, cost-effectiveness analysis, detection and discussion of biologically relevant factors, and external validation.

Conclusion: DLR has potential to improve model performance in HNC prognosis.

Keywords: deep learning, radiomics, PET/CT, head and neck cancer, prognosis

1. INTRODUCTION

HNC is the sixth most common type of cancer worldwide, with approximately 600,000 new cases each year, most of which are head and neck squamous cell carcinoma (HNSCC); its mortality rate reaches 40–50% [1]. The main risk factors for HNC include smoking, alcohol consumption, and infection with human papillomavirus [2].
Treatments include surgery, radiotherapy, and chemotherapy. Information on the shape, size, and metabolism of head and neck tumors is provided primarily by positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI) [3]. Although excellent progress has been made in treatment technology, and the staging systems are being updated, the recurrence rate of HNC in patients is as high as 40%, and the prognosis remains unsatisfactory [4]. Visual analysis allows for diagnosis, staging, and detection of recurrence, but is subjective. Quantitative analysis appears necessary for disease prognosis [5]. In the past decade, the conversion of standard medical images into analyzable high-dimensional data has been at the forefront of imaging research [6]. Traditional medical imaging analysis is subjective and qualitative. The subjectivity of radiologists significantly affects analysis results. Radiomics and deep learning for quantitative analysis have attracted increasing attention because they can avoid this subjectivity [7].

Information in medical images can be obtained through both visualization with the human eye and quantitative analysis. In recent years, radiomics for quantitative analysis has become a new and rapidly developing research area in medical imaging. Radiomics based on machine learning uses advanced techniques to quantitatively extract information with diagnostic, predictive treatment response, and prognostic value from medical images such as PET, CT, and MRI. The information is analyzed, and the results can be applied to clinical decision-making for more accurate diagnosis and prognosis [8-11]. Conventional radiomics workflows include ROI segmentation, feature extraction and selection, model construction, and result analysis (Figure 1). Radiomics is based on manually constructed features, such as histograms, texture and shape features, and machine learning classifiers [12]. Although the application of radiomics in head and neck tumor prognosis has become more sophisticated, traditional machine learning methods and the limited number of variables have resulted in some shortcomings of these models [13].

Conventional radiomics uses manual or semi-automated segmentation of tumor ROIs, extracts predefined features, and then applies them to machine learning classification models. In contrast, deep learning radiomics (DLR) applies deep learning in the steps of ROI segmentation, feature extraction, and model building, thus providing the advantages of both deep learning and conventional radiomics. Deep convolutional neural network (DCNN) segmentation models, particularly the U-Net architecture, has led to tremendous progress in segmenting ROIs [14]. Deep learning typically extracts deep features from the last or penultimate convolutional layer of a neural network, which are complementary to those of radiomics in some aspects; therefore, DLR has potential in cancer prognosis [15]. To solve complex computational problems, deep neural networks are more applicable than traditional machine learning approaches, such as logistic regression, owing to the multilayer neural network structure [16].

After studying the application of DLR to other diseases. Shboul et al. have used a feature-based random forest segmentation model and a convolutional neural network (CNN)-based featureless segmentation model to segment ROIs, followed by an “union” approach to combine the segmentation outputs. In the featureless segmentation described above, three network models, DCNN, U-Net, and Fully Connected Network, were used in combination [17]. Paul et al. have used three pre-trained CNNs (Vgg-F, Vgg-M, and Vgg-S) to extract features, which were later combined with traditional radiomics features after feature filtering, then fed into a classifier to classify benign malignancy in lung nodules [18]. Devinder et al. have extracted features directly by using a five-layer auto-encoder, and output feature vectors in the penultimate layer of the network, followed by use of binary classifiers to distinguish between benign from malignant lung nodules [19]. Similarly to Paul et al., Lao et al. have used a deep learning model...
that extracts deep features and output features in the penultimate layers, and combines them with other radiomics features to predict survival in patients with glioblastoma multiforme [20]. In contrast to the approach of Lao et al., Wang et al. have directly used CNN for feature extraction and applied a conventional classifier to predict the grading of low-grade glioma mutations [21]. Some studies have used DLR methods to predict the prognosis of HNC. This article systematically reviews the application of deep learning combined with conventional radiomics in the prediction of HNC prognosis.

2. MATERIALS AND METHODS

2.1 Eligibility criteria

2.1.1 Inclusion criteria. The inclusion criteria were as follows: (1) observational studies (e.g., prospective cohort studies, retrospective cohort studies, or case-control studies) or clinical trials (e.g., randomized controlled trials); (2) studies in populations of patients diagnosed with HNC; (3) PET, CT, or PET/CT-based patient prognosis studies; (3) studies with methods combining traditional imaging histology and deep learning; and (4) studies with complete data available.

2.1.2 Exclusion criteria. The following exclusion criteria were applied: (1) non-human studies; (2) studies lacking sufficient data; (3) studies that did not evaluate deep learning combined with conventional radiomics; (4) case reports, reviews, experimental studies, short communications, personal opinions, letters to the editor, or conference abstracts; and (5) non-English language studies.

2.2 Focus

This article summarizes and discusses the application of DLR in the prediction of HNC prognosis, and compares the performance of DLR with that of conventional radiomics alone.

2.3 Search strategy

Electronic databases (PubMed, Embase, Scopus, Web of Science, and Cochrane) were searched with the following terms: (head and neck neoplasms OR head and neck cancer AND machine learning OR deep learning OR artificial intelligence OR radiomics OR DLR). The search period was from 2012 through 2022.

2.4 Study selection and quality assessment

Of the 796 studies evaluated, a total of 788 were excluded, and eight were included (Figure 2). Two investigators (Bingzhen Wang and Yifan Yang) screened the studies.

The RQS, a radiomics-specific quality assessment tool, was used to assess the methodological quality of

Figure 2 | Flow diagram of literature search and selection.
the included studies [22]. Sixteen items were assessed, including imaging protocol, feature extraction, data modeling, model validation, and data sharing. The summed total score ranged from -8 to 36, and was converted into a final 0–100 percentage score [23]. The RQS scoring table is shown in Table 1. Two researchers evaluated the articles and reached a consensus.

3. RESULTS

3.1 Study selection and characteristics
Among the eight articles screened, all of which were retrospective studies, four were published in 2022, one was published in each of 2021 and 2020, and the remaining two were published in 2019. Data from the articles were extracted and analyzed. In total, 2339 patients were included in the studies. The included studies reached a median total RQS of 13 with a range of 10–15, corresponding to percentages of 36.11% and 27.78–41.67%. The specific literature score table is shown in Table 2. The included articles all used deep learning combined with conventional radiomics, and four studies detailed the toolkits used for feature extraction. Details on the toolkits, deep learning networks, and conventional radiomics methods used in the articles are shown in Table 3. The number of features extracted, feature selection methods, and final results are shown in Table 4.

3.2 Main findings
Traditional radiomics tended to manually or semi-automatically segment ROI, extract defined features, and apply machine learning models for prediction. Radiomics combined with deep learning enabled automatic segmentation of ROIs, extraction of deeper measured features, and building of deep network models. Of the eight included articles, three used deep learning for automatic segmentation of ROIs [24-26], five combined deep learning in feature extraction [26-29] and feature fusion [30], and two used deep learning for model building [27, 31].

3.2.1 Deep learning for ROI segmentation
In terms of ROI segmentation, automatic segmentation using deep learning eliminated the subjective influence of manual segmentation and enabled more accurate ROI segmentation, through use of both low and high-level details, thus enabling extraction of more accurate and comprehensive features, and improving model predictive performance [24-26]. Salamanpour et al. used 3D U-Net and 3D U-NETR to automatically segment images of HNC. Five fusion techniques, seven dimensionality reduction algorithms, and five survival prediction algorithms were used to predict the survival of patients. The Laplacian pyramid + spare representation fusion technique with the 3D U-Net model obtained the highest segmentation ACC, with a Dice score of 0.81 on the validation set, because 3D U-Net had more layers and a similar number of parameters. On the basis of the optimal fusion and segmentation, 215 radiomics features were extracted from PET and CT images with the Standardized Environment for Radiomics Analysis (SERA) package. Moreover, the Gradient Boosting with Component-wise Linear Models (GlmBoost) survival prediction algorithm was used to predict survival. Finally the Ensemble Voting technique was used to predict the survival rate, with a C-index of 0.75 on the validation set [24]. Mehdi et al. have also used multiple fusion techniques to combine PET/CT images and used deep learning for automatic segmentation of ROI. The Laplacian Pyramid fusion technique combined with 3D U-Net had the best segmentation performance, with a Dice score of 0.76 in the validation set. A total of 215 radiomics features were extracted by the SERA imaging histology package and input to multiple hybrid machine learning systems containing 13 dimensionality reduction algorithms, 15 feature selection algorithms, and 8 survival prediction algorithms. The final prediction results were obtained with a C-index value of 0.73 on the validation set [25] by using the Ensemble Voting technique. Bourigault et al. have used an encoder-decoder combining 3D Normalised Squeeze-and-Excitation Residual Blocks proposed by lantsen with the full-size connected U-Net's UNet3 + structure proposed by Huang [32, 33]. The network structure was capable of downward and upward sampling, focused on the image ROI as well as other relevant regions; combined local and global full-scale information; and leveraged low-level and high-level semantics, thus improving the ACC of segmentation. An average Dice similarity coefficient (DSC) of 0.75 was obtained in the cross-validation [26]. The tumor segmentation with the deep learning approach achieved high ACC segmentation and laid a good foundation for subsequent steps.

3.2.2 Deep learning radiomics for feature extraction and feature fusion
In terms of feature extraction and feature fusion, DLR automatically learned features from images, which differed from the features manually extracted through radiomics. Because the two classes of features above obtained complementary information, deep learning features and radiomics features can be combined to produce more stable prediction results [26-29].

Chen et al. input PET and CT images into a multi-objective radiomics (MaO-Radiomics) model with support vector machine (SVM) classifiers and a 3D-CNN model consisting of convolutional layers, max-pooling layers, and fully connected layers, which fully used contextual information to extract features and made predictions, respectively. The outputs were finally combined with an evidential inference evidential reason (ER) method. The hybrid model combining the DCNN model and the MaO-Radiomics model had an ACC of 0.88 for the classification of normal, suspicious, and involved LNs, thus indicating better performance than the ACC values of 0.75 obtained with conventional radiomics alone.
| Criteria                                                                 | Points                                                                 |
|------------------------------------------------------------------------|------------------------------------------------------------------------|
| **Item 1** Image protocol quality: well documented image protocols (e.g., contrast, slice thickness, energy, etc.) and/or use of public image protocols enable reproducibility/replicability. | +1 (if protocols are well documented) +1 (if public protocol is used) |
| **Item 2** Multiple segmentations: possible actions include segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, and segmentation at different breathing cycles. Analysis of feature robustness to segmentation variability. | +1 |
| **Item 3** Multiple segmentations: possible actions include segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, and segmentation at different breathing cycles. Analysis of feature robustness to segmentation variability. | +1 |
| **Item 4** Imaging at multiple time points: collection of individuals' images at additional time points. Analysis of feature robustness to temporal variability (e.g., organ movement, organ expansion/shrinkage). | +1 |
| **Item 5** Feature reduction or adjustment for multiple testing decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consideration of feature robustness when selecting features. | -3 (if neither measure is implemented) +3 (if either measure is implemented) |
| **Item 6** Multivariable analysis with non-radiomic features (e.g., EGFR mutation) is expected to provide a more holistic model permitting correlations/inferences between radiomics and non-radiomics features. | +1 |
| **Item 7** Detecting and discussing biological correlates: demonstration of phenotypic differences (possibly associated with underlying gene–protein expression patterns) deepens understanding of radiomics and biology. | +1 |
| **Item 8** Cut-off analyses: determination of risk groups by either the median or a previously published cut-off, or report a continuous risk variable. Decreases the risk of reporting overly optimistic results. | +1 |
| **Item 9** Discrimination statistics: reporting of discrimination statistics (e.g., C-statistic, ROC curve, or AUC) and their statistical significance (e.g., p-values or confidence intervals). A resampling method (e.g., bootstrapping or cross-validation) can also be applied. | +1 (if a discrimination statistic and its statistical significance are reported) +1 (if a resampling method technique is also applied) |
| **Item 10** Calibration statistics: reporting of calibration statistics (e.g., calibration-in-the-large/slope or calibration plots) and their statistical significance (e.g., p-values or confidence intervals). A resampling method (e.g., bootstrapping or cross-validation) can also be applied. | +1 (if a calibration statistic and its statistical significance are reported) +1 (if a resampling method technique is also applied) |
| **Item 11** A prospective study registered in a trial database provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker. | +7 (for prospective validation of a radiomics signature in an appropriate trial) |
Table 1 | Continued

| Criteria                                                                 | Points |
|--------------------------------------------------------------------------|--------|
| Validation performed without retraining and adaptation of the cut-off value, thus providing crucial information concerning credible clinical performance. | -5 (if validation is missing) +2 (if validation is based on a dataset from the same institute) +3 (if validation is based on a dataset from another institute) +4 (if validation is based on two datasets from two distinct institutes) +4 (if the study validates a previously published signature) +5 (if validation is based on three or more datasets from distinct institutes) |
| Comparison to “gold standard”: assessment of the extent to which the model agrees with or is superior to the current “gold standard” method” (e.g., TNM staging for survival prediction). This comparison indicates the added value of radiomics. | +2 |
| Potential clinical utility: reporting on the current and potential application of the model in a clinical setting (e.g., decision curve analysis). | +2 |
| Cost-effectiveness analysis: reporting on the cost-effectiveness of the clinical application (e.g., quality adjusted life years generated). | +1 |
| Open science and data: code and data made publicly available. Open science facilitates knowledge transfer and reproducibility of the study. | +1 (if scans are open source) +1 (if region of interest segmentations are open source) +1 (if code is open source) +1 (if radiomics features are calculated on a set of representative ROIs and the calculated features + representative ROIs are open source) |

Total points (36=100%)
Bizzego et al. observed that the approach of mixing DLR features has more accurate predictive performance than models using only one feature type or image pattern. Bizzego's study consisted of two identical parallel deep learning CNN networks trained on CT and PET simultaneously (including BatchNorm layers, convolutional layers, Dropout layers, MaxPool3d layers, and an AdaptiveAvgPool3D layer) to extract deep learning features output by the final AdaptiveAvgPool3D layer, and used a predefined feature extractor to extract radiomics features. Subsequently, 239 deep learning features and 261 radiomics features were concatenated and combined into a unified classification pipeline with linear SVM for prognosis of local recurrence of HNC. The ACC of hand-crafted radiomics (HCR) + DLR was 0.96 on the test set, a value higher than the ACC associated with using only manual or deep learning features (ACC of 0.87 and 0.79, respectively) [28].

Peng et al. used ITK-SNAP to manually segment PET and CT images. Four DNNs (12 or 8 weighted layers) were constructed to extract deep learning features. One of the DNNs for CT images had three convolutional layer groups, followed by a dropout layer and a bidirectional join layer with a softmax classifier. Each convolutional layer group consisted of four convolutional layers and one max-pooling layer. Only two convolutional groups existed for PET images. Deep learning features were output by the last convolutional layer. Finally, 14 deep learning features and four radiomics features were identified, and linear combinations were weighted by their coefficients. The optimal combination of prognostic features was screened with Least absolute shrinkage and selection operator (LASSO) Cox regression, and the features were input into the Cox proportional risk model to construct a radiomics nomogram for predicting disease-free survival in HNC. A C-index of 0.722 was obtained in the test set, thus indicating a significant improvement over the C-index values of 0.634 and 0.655 obtained in the team's previous radiomics studies using manual features based on CT and PET [34-37]. The combination of manual features and depth features had the best discriminatory ability [29].

Bourigault et al. used automatic segmentation of ROI and 3D U-Net for deep feature extraction from the fifth convolutional layer, the PyRadiomics package for extracting radiomics features, and 5-fold cross-validated LASSO regression for feature selection. They finally identified 49 deep learning features and 14 radiomics features, combined with 7 clinical features in a Cox proportional risk regression. The model was subsequently used to predict progression-free survival of patients and exhibited a C-index of 0.82 on the validation set, thereby indicating a significant improvement over the C-index of 0.72 obtained with only radiomics features; however, the C-index of 0.62 obtained on the external challenge test set exhibited an overfitting
### Table 3 | Deep learning and radiomics models.

| Author (year)            | Number of patients | Segmentation method | Whether to use deep learning | Package                                      | Deep learning networks                                                                 | Conventional radiomics methods         |
|--------------------------|--------------------|---------------------|-----------------------------|----------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------|
| Chen et al. (2019)       | 59                 | Manual              | Yes                         | NR                                           | 3D-CNN (12 convolutional layers, 2 max-pooling layers, and 2 fully connected layers)     | SVM                                    |
|                          |                    |                     |                             |                                              | Each convolutional layer is equipped with ReLU activation and batch normalization         |                                        |
| Bizzego et al. (2019)    | 194                | Manual              | Yes                         | NR                                           | 3D multimodal CNN                                                                         | LSVM                                   |
| Salmanpour et al. (2022) | 325                | Auto                | Yes                         | SERA package for radiomics features          | 3D U-NET                                                                                 | CSF; CoxPH; FSSVM; CoxBoost; GlmBoost; GBM; RSF |
| Mehdi et al. (2022)      | 325                | Auto                | Yes                         | SERA package for radiomics features          | 3D U-NET; SegResNet                                                                      | CSF; CoxPH; FSSVM; CoxBoost; GlmBoost; GBM |
| Peng et al. (2022)       | 707                | Manual              | Yes                         | Python Keras package with the Tensorflow library | DCNNs (12 or 8 weighted layers)                                                            | Cox regression                         |
| Tang et al. (2021)       | 188                | Manual              | Yes                         | NR                                           | DL-ANN (3 hidden layers, 107 inputs, and 1 binary output)                                | 3D slicer (v. 4.10.2) with PyRadiomics |
| Zhou et al. (2020)       | 188                | Manual              | Yes                         | NR                                           | DL with stacked sparse autoencoder                                                        | SVM; DT; KNN                           |
| Bourigault et al. (2022) | 353                | Auto                | Yes                         | PyRadiomics package for radiomics features   | 3D NormResSE-Unet3 + (an encoder-decoder architecture with full inter- and intra-skip connections) | CoxPH regression model                  |

CNN: convolutional neural network; DL: Deep learning; SVM: support vector machine; LSVM: linear support vector machine; CSF: conditional survival forest; CoxPH: Cox’s proportional hazard; FSSVM: Fast Survival SVM; CoxBoost: CoxPH model by likelihood-based Boosting; GlmNet: LASSO and Elastic-Net Regularized Generalized Linear Models; GlmBoost: Gradient Boosting with Component-wise Linear Models; GBM: Gradient boosting machines; RSF: random survival forest; DT: decision tree; KNN: K-nearest neighbor; NR: not reported
problem that could be decreased by adding regularization to the model [26].

In terms of feature fusion, different imaging modalities measured different features. For example, PET scans measured glucose metabolism, and CT scans provided attenuation coefficient information. When features came from different sources and were complementary, feature fusion was required. A stacked self-encoder depth model was used to combine features from different modalities into one feature set, instead of simply stitching multimodal features into one long vector, as in conventional radiomics, to improve the predictive performance of the model. Zhou et al. extracted radiomics features from manually segmented ROIs; used a stacked sparse auto-encoder structure to combine 257 manual features obtained from PET and CT images; and input them into a MoA-radiomics model with multiple basis classifiers such as SVM, decision tree (DT), and K-nearest neighbor for prognostic prediction of HNC with distal metastasis, rather than simply concatenating features extracted from different modalities. Evidence-based reasoning was used to combine the outputs of multiple models at the decision level, thus resulting in an AUC of 0.84, indicating a significant improvement over the results of feature fusion without using stacked self-encoders. The improvement was attributed to the stacked self-encoder’s extracting more distinguishing features and discovering joint information among different patterns [30].

### 3.2.3 Deep learning for model building.

In model building, deep learning combined with radiomics has higher predictive performance than radiomics alone [27, 30]. Chen et al. used the ER method to combine the previously introduced deep learning-based 3D-CNN and MoA-Radiomics models, in a simple, easily implemented process. The input image was normalized to accelerate the convergence, and the synthetic minority over-sampling technique was used to balance and expand the minority class samples to improve the efficiency of the model. The final ACC was 0.88, indicating an improvement over radiomics alone (ACC of 0.75) and MoA-Radiomics alone (ACC of 0.82) [27]. Tang et al. used 3D slicer with PyRadiomics extensions to extract 107 radiomics features from CT images and input them into an artificial neural network with hidden layers, then used the binary output to predict death prognosis and cancer recurrence rates. The imaging data were used as both the training and validation sets. The AUC for cancer recurrence with gross tumor volume in this experiment was 0.956, and the ACC was 0.724 [30]. Artificial neural networks outperformed other methods for most data predictions, and performed well with higher data complexity [38].
Table 5 | Comparison table of evaluation indicators.

| Author (year) | Methods or features | Evaluation metrics |
|---------------|---------------------|--------------------|
| Chen et al. (2019) | Radiomics          | ACC=0.75           |
|                | MaO-radiomics       | ACC=0.82           |
|                | DCNN + MaO-radiomics| ACC=0.88           |
| Bizzego et al. (2019) | HCR               | ACC=0.79           |
|                | DLR                | ACC=0.87           |
|                | HCR + DLR          | ACC=0.96           |
| Peng et al. (2022) | CT-HCR             | C-index=0.63       |
|                | PET-HCR            | C-index=0.65       |
|                | HCR + DLR          | C-index=0.72       |
| Bourigault et al. (2022) | Clinical + PET radiomics | C-index=0.67 |
|                | Clinical + CT radiomics | C-index=0.68 |
|                | Clinical + PET/CT radiomics | C-index=0.72 |
|                | Clinical + CT radiomics + deep learning features | C-index=0.82 |

The performance data from the included studies involving DLR, compared with radiomics, are shown in Table 5.

4. DISCUSSION

In this systematic review, we evaluated the literature on the application of radiomics combined with deep learning in the prognosis of HNC. In recent years, the application of DLR for the quantitative analysis of medical images has increased significantly [39-41]. Radiomics efforts have focused on manually extracting radiomics features, such as texture features and histogram features. Feature filtering is used to filter the optimal set of features, and the features are then fed into machine learning classifiers (e.g., SVM or DT) [7]. Several radiomics studies on HNC have been reported. For example, Wang et al. [42] have used radiomics combined with a machine learning model to predict the T stage of locally advanced laryngeal cancer. Ren et al. [43] have extracted MRI-based radiomics features to stage HNSCC III-IV and I-II. Yuan et al. [44] have found that MRI-based radiomics features are independent prognostic factors in patients with HNSCC. Other studies have combined clinico pathological features with radiomics features to predict overall survival and disease-free survival [45, 46] or have linked radiomics findings to molecular features of HNC [47-49]. However, manually mapping ROIs is time consuming and characterized by high inter-rater variability. Deep learning can solve this problem through automatic ROI segmentation, and it has also shown great potential in extracting features and combining fully connected layers to accomplish classification and prediction.

The three articles on automatic segmentation reviewed in this study all used 3D U-Net for automatic segmentation. Because the U-Net network is relatively simple and combines down sampling and up sampling paths, and has the ability to aggregate features and spatial information for accurate localization, it is commonly used for the segmentation of medical maps. For example, Lin et al. [50] have used it for automatic segmentation of cervical cancer, and Moe et al. [51] and Ren et al. [52] have used it for delineation of gross tumor volume.

DLR and HCR can be combined in two approaches: decision-level and feature-level fusion [56]. In decision-level fusion, separate classifiers are trained by using DLR and HCR, and then the results are combined with ER to obtain the final classification. Chen et al. [27] have combined CNN and radiomics results by using ER. This process can also be performed by using voting, such as soft voting. In feature-level fusion, DLR and HCR are concatenated into feature vectors, and then the classifier is trained, as reported by Bizzego et al. [28]

Important components of deep learning models include convolutional and fully connected layers, which can automatically learn features and therefore be used to extract deep features of data. In the four articles extracting deep learning features included in this study, the convolutional layers ranged between 5 and 12 layers, to avoid overfitting due to overly complex parameters. Bourigault et al. used the output of the fifth convolutional layer of the 3D-UNet network as a deep feature, Bizzego et al. used the output of the AdaptiveAvgPool3d layer after the last convolutional...
layer as a deep feature, and the other studies used the output from the last convolutional layer as a deep feature. Specific information on the network was presented in the results section. Deep learning networks as feature extractors have also been applied in other cancer studies, such as Lao et al. [20] for survival prediction for patients with glioblastoma multiforme; Wang et al. [66] for Coronavirus disease 2019 pneumonia, and Wang et al. [67] for hilar cholangiocarcinoma.

CNN, a class of deep learning models, can learn deeper features from images and eventually map the input image to the desired output for prediction. In the two studies of radiomics combined with deep learning models reviewed herein, Chen et al. and Tang et al. predicted the results with deep networks, whereas Chen et al. combined CNN results and radiomics model results at the decision level. Some studies have used deep learning models in HNC, such as Lombardo et al. [68] on long-range metastasis in HNC, Kann et al. [69] on nodal metastasis and extranodal extension, and Naser et al. on progression free survival prediction; moreover, Kim et al. [70] built a deep learning-based survival prediction algorithm named DeepSurv for survival prediction among patients with oral cancer. Notably, large amounts of data are needed to train deep learning networks to obtain better performance.

Most articles included in our study were based on PET/CT imaging data, wherein images based on different modalities contained different tumor information. Most current radiomics studies have combined the features of multiple modalities only in a simple linear way and thus have not benefited from the advantages of multimodality. Integration of the features of multiple modalities organically in various ways, such as using stacked sparse auto-coding, can increase model accuracy [71, 72]. ER methods have been used to combine the prediction results of multiple classifiers, or to combine the results of radiomics and deep learning, to improve prediction results [30]. Overfitting occurs when the number of extracted features is larger than the number of samples, and can be avoided by adding regularization or using a feature filtering algorithm to obtain the optimal set of features. Multiple algorithms can be combined instead of use of a single algorithm for pre-experimentation, and the best algorithm can be applied to filter the features.

In conclusion, the application of combined DLR improves HNC prognosis beyond that achievable with conventional radiomics. However, deep learning also has limitations, and some problems remain to be improved. First, deep learning networks contain many parameters, and a massive number of high-quality samples are needed to train the network and avoid overfitting. Small sample size was a common problem and was the main limitation of model performance. When the training samples are not sufficiently large, transfer learning pre-trained networks or data augmentation techniques can be used to expand the sample data, such as random addition of noise, rotation or inversion of images, etc., to further improve model performance. Bourigault et al. randomly flipped the tumor volume left/right, superior/inferior, and anterior/posterior in their study to enhance the data [26]. Chen et al. performed a data enhancement technique based on 59 patients, by adding samples from both suspicious and normal categories to balance the data [27]. Bizzegio et al. pre-trained the architecture for T-stage then minimal rotation, translation and used Gaussian noise to adjust the images [28]. Moreover, in Tang et al., all data were used as a training cohort and validation cohort simultaneously, by leaving one set as the validation set and using the remainder for training. The above process was repeated until all data were used for the validation set. This method is also effective in training and validation for small sample sizes [31]. Unsupervised learning does not require tags, for example using Deep Autoencoders and The Restricted Boltzmann Machine, etc. Chang et al. proposed a multi-scale convolutional sparse coding method to provide an unsupervised solution [73]. In the future, semi-supervised learning could also be applied to self-train a small amount of labeled data, and unlabeled data could then be studied to solve the problem of a lack of samples being labeled [74, 75]. Second, the features extracted by radiomics and deep learning, respectively, came from the same imaging modality, and redundancy in the feature space negatively affected model performance. Chen et al. [27] and Zhou et al. [30] used ER for combining the probabilistic outputs of multiple models at the decision level to decrease the influence of feature space redundancy on model performance. Third, owing to their different principles, radiomics and deep learning have different advantages for specific tasks. Deep neural networks focus on the whole image, and are preferred in image fusion, feature extraction, and tumor segmentation tasks [76]. Although two studies on predictive model building by deep neural networks included herein had good performance, training deep learning models requires large amounts of data. When the amount of image data is small, deep neural networks cannot fully outperform traditional machine learning classifiers in model building [77]. However, deep neural networks still have great potential in prediction tasks. With the rapid development of deep learning technology, the use of deep learning must be explored to build prediction models in future research. Finally, the stability of deep learning features also must be studied. Compared with conventional radiomics, black box deep learning methods are less interpretable, and the features learned by deep learning are complex to interpret and conceptualize.

In addition, some limitations exist in the current radiomics studies. First, most studies were based on retrospective single-center data, but models require external validation with multi-center data to improve generalizability and ACC [78]. Therefore, multi-center prospective studies and clinical trials are necessary [79] to determine whether DLR might be applied in clinical settings. The data in Peng et al. were acquired from one center, and
prospective studies with external validation to verify their results were lacking [29]. Although some data platforms have been established, such as The Cancer Impact Archive, the quality of the data varies. By establishing a unified standard to ensure data quality also must be addressed as soon as possible [80]. Second, combining clinical features (e.g., age, sex, TNM stage, etc.) could make the model more comprehensive [81]. Clinical features were included in studies such as Salmanpour et al. [24], Mehdi et al. [25], and Chen et al. [27]. Consequently, appropriate clinical features should be added when constructing features to improve model performance.

5. CONCLUSION

PET/CT-based DLR has promising prospects in HNC prognosis. Among them, deep learning can be used for ROI segmentation, feature extraction and fusion, and model building. In ROI segmentation, automatic segmentation can avoid the subjective influence of manual segmentation, and combine global and local information to make the segmentation of ROI more accurate. In feature extraction and fusion, images of different modalities include different features, and conventional radiomics features and deep learning features are complementary. Combining deep features with radiomics features and using deep learning to combine features from different imaging modalities greatly improves overall model predictive performance. In model building, combining deep learning models results in faster prediction times and good performance in handling relatively complex datasets. Accordingly, adding deep learning techniques to conventional radiomics results in higher evaluation metrics in HNC prognosis than using only conventional radiomics. Of course, the prediction results are also affected by the number of cases, centers, and the quality of data, thus potentially leading to overfitting and other problems. In short, the organic combination of both methods improves model performance.

SOURCE OF FUNDING

This work was supported by National Natural Science Foundation of China (62276022, 61876165, 61503326), Hebei Province Introduced Returned Overseas Chinese Scholars Funding Project (C20220107), Hebei Natural Science Foundation (C2022406010) and Technology Innovation Guidance Project-Science and Technology Work Conference.

REFERENCES

[1] Cognetti DM, Weber RS, Lai SY. Head and neck cancer: an evolving treatment paradigm. Cancer 2008;113:1911-32. [PMID: 18798532; DOI: 10.1002/cncr.23654]
[2] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBCAN 2012. Int J Cancer 2015;136:E359-86. [PMID: 25220842; DOI: 10.1002/ijc.29210]
[3] Butowsk NA. Epidemiology and diagnosis of brain tumors. Continuum 2015;21:301-13. [PMID: 25837897; DOI: 10.1212/201.CO1.0000464171.50638.fa]
[4] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424. [PMID: 30207593; DOI: 10.3322/caac.21492]
[5] Castelli J, De Bari B, Depeursinge A, Simon A, Devillers A, et al. Overview of the predictive value of quantitative 18 FDG PET in head and neck cancer treated with chemoradiotherapy. Crit Rev Oncol Hematol 2016;108:40-51. [PMID: 27931839; DOI: 10.1016/j.critrevonc.2016.10.009]
[6] Sanduleanu S, Woodruff HC, de Jong ECC, van Timmeren JE, Jochems A, et al. Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. Radiother Oncol 2018;127:349-60. [PMID: 29779918; DOI: 10.1016/j.radonc.2018.03.033]
[7] Peng Z, Wang Y, Wang Y, Jiang S, Fan R, et al. Application of radiomics and machine learning in head and neck cancers. Int J Biol Sci 2021;17:475-86. [PMID: 33613106; DOI: 10.7150/ijbbs.55716]
[8] Gardin I, Grégoire V, Gibon D, Kirilsi H, Pasquier D, et al. Radiomics: principles and radiotherapy applications. Crit Rev Oncol Hematol 2019;138:44-50. [PMID: 31092384; DOI: 10.1016/j.critrevonc.2019.03.015]
[9] Heukelom J, Fuller CD. Head and neck cancer Adaptive Radiation Therapy (ART): conceptual considerations for the informed clinician. Semin Radiat Oncol 2019;29:258-73. [PMID: 31027643; DOI: 10.1016/j.semradiol.2019.02.008]
[10] Mayerhoefer ME, Materka A, Langs G, Hägström I, Szczypiński P, et al. Introduction to radiomics. J Nucl Med 2020;61:488-95. [PMID: 32060219; DOI: 10.2967/jnumed.118.222893]
[11] Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. Radiology 2021;298:505-16. [PMID: 33399513; DOI: 10.1148/radiol.212102553]
[12] Rizzo S, Botta F, Raimondo S, Orliggi D, Fanciullo C, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018;2:36. [PMID: 30426318; DOI: 10.1186/s41747-018-0068-z]
[13] Limkin EJ, Sun R, Dercole L, Zacharaki EI, Robert C, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol 2017;28:1191-206. [PMID: 28168275; DOI: 10.1093/annonc/mdx034]
[14] Havaei M, Davy A, Warde-Farley D, Biard A, Courtville A, et al. Brain tumor segmentation with deep neural networks. Med Image Anal 2017;35:18-31. [PMID: 23399513; DOI: 10.1016/j.medima.2016.05.004]
[15] Dong D, Fang M-J, Tang L, Shan X-H, Gao J-B, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann Oncol 2020;31:912-20. [PMID: 32304748; DOI: 10.1016/j.annonc.2020.04.003]
[16] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, et al. Deep learning in cancer diagnosis, prognosis and treatment selection. Frontiers in Oncology 2019;11:964. [PMID: 31619949; DOI: 10.3389/fonci.2019.00966]
[18] Paul R, Hawkins SH, Schabath MB, Gillies RJ, Hall LO, et al. Predicting malignant nodules by fusing deep features with classical radiomics features. J Med Imaging (Bellingham) 2018;5:011021. [PMID: 29594181; DOI: 10.1117/1.JMI.5.1.011021]

[19] Kumar D, Wong A, Clausi DA. Lung nodule classification using deep features in CT images. In: 2015 12th Conference on Computer and Robot Vision; 2015.

[20] Lao J, Chen Y, Li Z-C, Li Q, Zhang J, et al. A deep learning-based radiomics model for prediction of survival in Glioblastoma Multiforme. Sci Rep 2017;7:10353. [PMID: 28871110; DOI: 10.1038/s41598-017-10649-8]

[21] Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep 2017;7:5467. [PMID: 28710497; DOI: 10.1038/s41598-017-05848-2]

[22] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-62. [PMID: 28975929; DOI: 10.1038/nrclinonc.2017.141]

[23] Ponsiglione A, Stanzione R, Ascione R, Gambardella M, et al. Cardiac CT and MRI radiomics: systematic review of the literature and radiomics quality score assessment. Eur Radiol 2022;32:2629-38. [PMID: 34812912; DOI: 10.1007/s00330-021-08375-x]

[24] Salmanpour MR, Hajianfar G, Rezaeijo SM, Ghaemi M, Rahmim A, Folkest M, Liu H, Jiang S, et al. Multifaceted radiomics for distant metastasis prediction in head & neck cancer. Phys Med Biol 2020;65:155009. [PMID: 32294632; DOI: 10.1088/1361-6560/ab8956]

[25] Fatam M, Hosseinzadeh M, Askari D, Shelikh H, Rezaeijo SM, et al. Fusion-based head and neck tumor segmentation and survival prediction using robust deep learning techniques and advanced hybrid machine learning systems. In: Head and neck tumor segmentation and outcome prediction. vol 13209. Cham: Springer; 2022:177-86.

[26] Bourigault E, McGowan DR, Mehranian A, Papiez BW. Multimodal PET/CT tumour segmentation and prediction of progression-free survival using a full-scale U-Net with attention. In: Head and neck tumor segmentation and outcome prediction. Cham: Springer International Publishing; 2022.

[27] Chen L, Zhou Z, Sher D, Zhang Q, Shah J, et al. Combining many-objective radiomics and 3D convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer. Phys Med Biol 2019;64:075011. [PMID: 30780137 DOI: 10.1088/1361-6560/ab083a]

[28] Bizzego A, Bussola N, Salvalai D, Chierici M, Maggio V, et al. Integrating deep and radiomics features in cancer biomaging. In: 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2019;1-8. [DOI: 10.1109/CIBCB.2019.8791473]

[29] Peng H, Dong D, Fang M-J, Li L, Tang L-L, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal Carcinoma. Clin Cancer Res 2019;25:4271-9. [PMID: 30975664; DOI: 10.1158/1078-0432.CCR-18-3065]

[30] Zhou Z, Wang K, Folkest M, Liu H, Jiang S, et al. Radiomics AI prediction for head and neck squamous cell carcinoma (HNSCC) prognosis and recurrence with target volume approach. BJR Open 2021;3. [PMID: 34381946; DOI: 10.1259/bjro.20200073]

[31] Iantsen A, Visvikis D, Hatt M. Squeeze-and-excitation normalization for automated delineation of head and neck primary tumors in combined PET and CT images. In: Head and Neck Tumor Segmentation. Cham: Springer International Publishing; 2021.

[32] Huang H, Lin L, Tong R, Hu H, Zhang Q, et al. UNet 3+: a full-scale connected U-Net for medical image segmentation. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2020:1055-9.

[33] Huang Y, Liu Z, He L, Chen X, Pan D, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in Early-Stage (I or II) non-small cell lung cancer. Radiology 2016;281:947-57. [PMID: 27347764; DOI: 10.1148/radiol.2016152234]

[34] Huang YQ, Liang C-H, He L, Tian J, Liang C-S, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 2016;34:2157-64. [PMID: 27138577; DOI: 10.1200/JCO.2015.65.9128]

[35] Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. Radiology 2016;281:382-91. [PMID: 27144536; DOI: 10.1148/radiol.2016152110]

[36] Zhang B, Tian J, Dong D, Gu D, Dong Y, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 2017;23:4259-69. [PMID: 28280088; DOI: 10.1158/1078-0432.CCR-16-2910]

[37] Soudy M, Alam A, Ola O. Predicting the cancer recurrence using artificial neural networks. In: Raza K, editor. Computational Intelligence in Oncology: Applications in Diagnosis, Prognosis and Therapeutics of Cancers. Singapore: Springer; 2022:177-86.

[38] Muhl Bauer J, Egen L, Kowalewski K-F, Grilli M, Walach MT, et al. Radiomics in renal cell carcinoma—a systematic review and meta-analysis. Cancers (Basel) 2021;13:1348. [PMID: 33802699; DOI: 10.3390/cancers13061348]

[39] Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, et al. A roadmap for foundational research on Artificial Intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The academy workshop. Radiology 2019;291:781-91. [PMID: 30990384; DOI: 10.1148/radiol.2019190613]

[40] Choy G, Khalizadeh O, Michalski M, Do S, Samir AE, et al. Current applications and future impact of machine learning in radiology. Radiology 2018;288:318-28. [DOI: 10.1148/radiol.2018171820]

[41] Wang F, Zhang B, Wu X, Liu L, Fang J, et al. Radiomic nomogram improves preoperative T category accuracy in locally advanced laryngeal carcinoma. Front Oncol 2019;9:1064. [PMID: 31681598; DOI: 10.3389/fonc.2019.01064]

[42] Chen L, Tian J, Yuan Y, Dong D, Li X, et al. Magnetic resonance imaging based radiomics signature for the
preoperative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. Eur J Radiol 2018;106:1-6. [PMID: 30150029; DOI: 10.1016/j.ejrad.2018.07.002]

[44] Yuan, Y, Ren J, Shi Y, Tao X, Yuan Y, et al. MRI-based radiomic signature as predictive marker for patients with head and neck squamous cell carcinoma. Eur J Radiol 2019;117:193-98. [PMID: 31307647; DOI: 10.1016/j.ejrad.2019.06.019]

[45] Agarwal JP, Sinha S, Goda JS, Yoshikawa K, Mhatre R, et al. Tumor radiomic features complement clinicoradiological factors in predicting long-term local control and laryngectomy-free survival in locally advanced laryngeo- pharyngeal cancers. Br J Radiol 2020;93:20190857. [PMID: 32101463; DOI: 10.1259/bjr.20190857]

[46] Liu Z, Cao Y, Diao W, Cheng Y, Jia Z, et al. Radiomics-based prediction of survival in patients with head and neck squamous cell carcinoma based on pre- and post-treatment (18)F-PET/CT. Aging (Albany NY) 2020;12:14593-619. [PMID: 32674074; DOI: 10.18632/aging.103508]

[47] Zvirner K, Hilke FJ, Demidov G, Socarras Fernandez J, Ossowski S, et al. Radiogenomics in head and neck cancer: correlation of radiomic heterogeneity and somatic mutations in TP53, FAT1 and KMT2D. Strahlenther Onkol 2019;195:771-9. [PMID: 31123786; DOI: 10.1007/s00066-019-01478-x]

[48] Huang, C, Cintra M, Brenkan K, Zhou M, Colevas AD, et al. Development and validation of radiomic signatures of head and neck squamous cell carcinoma molecular features and subtypes. EBioMedicine 2019;49:70-80. [PMID: 31255659; DOI: 10.1016/j.ebiom.2019.06.034]

[49] Yoon, JH, Han K, Lee E, Lee J, Kim EK, et al. Radiomics in predicting mutation status for thyroid cancer: a preliminary study using radiomics features for predicting BRAFV600E mutations in papillary thyroid carcinoma. PLoS One 2020;15:e0228968. [PMID: 32053670; DOI: 10.1371/journal.pone.0228968]

[50] Lin YC, Lin CH, Lu HY, Chiang HJ, Wang HK, et al. Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer. Eur Radiol 2020;30:1297-305. [PMID: 31712961; DOI: 10.1007/s00330-019-06467-3]

[51] Moe YM, Groendahl AR, Tomic O, Dale E, Malinen E, et al. Deep learning-based auto-delineation of gross tumour volumes and involved nodes in PET/CT images of head and neck cancer patients. Eur J Nucl Med Mol Imaging 2021;48:2782-92. [PMID: 33559711; DOI: 10.1007/s00259-020-05125-x]

[52] Ren J, Eriksen JG, Nijkamp J, Korreman SS. Comparing different CT, PET and MRI multi-modality image combinations for deep learning-based head and neck tumor segmentation. Acta Oncol 2021;60:1399-406. [PMID: 34264157; DOI: 10.1007/s00432-021-02458-1]

[53] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-6. [PMID: 22257792; DOI: 10.1016/j.ejca.2011.11.036]

[54] Yamashita R, Nishio M, Do RKG, Ogaki K. Convolutional neural networks: an overview and application in radiology. Insights Imaging 2018;9:611-29. [PMID: 29934920; DOI: 10.1007/s13244-018-0639-9]

[55] Liu Z, Wang S, Dong D, Wei J, Fang C, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics 2019;9:1303-22. [PMID: 30867832; DOI: 10.7150/thno.30309]

[56] Afshar P, Mohammadi A, Plataniotis K, Oikonoumou A, Benali H. From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Processing Mag 2019;36:132-60. [DOI: 10.1109/MSP.2019.2900993]

[57] Wu G, Jochems A, Refaeet I, Ibrahim A, Yan C, et al. Structural and functional radiomics for lung cancer. Eur J Nucl Med Mol Imaging 2021 48(12);3961-74. [PMID: 33693966; DOI: 10.1007/s00259-021-05242-1]

[58] Wang X, Zhang L, Yang X, Tang L, Zhao J, et al. Deep learning combined with radiomics may optimize the prediction in differentiating high-grade lung adenocarcinomas in ground glass opacity lesions on CT scans. Eur J Radiol 2020;129:109150. [PMID: 32604042; DOI: 10.1016/j.ejrad.2020.109150]

[59] Afshar P, Mohammadi A, Tyrrell PN, Cheung P, Sigiu A, et al. (Formula: see text): deep learning-based radiomics for the time-to-event outcome prediction in lung cancer. Sci Rep 2020;10:12366. [PMID: 32703973; DOI: 10.1038/s41598-020-69106-8]

[60] Anastasaki A, Yang G, Zakko Y, Toma-Dasu I, Smedby O, et al. A comparative study of radiomics and deep-learning based methods for pulmonary nodule malignancy prediction in low dose CT images. Front Oncol 2021;11:737368. [PMID: 34976794; DOI: 10.3389/fonc.2021.737368]

[61] Liang HY, Yang SF, Zou H, Hou F, Duan LS, et al. Deep learning radiomics nomogram to predict lung metastasis in soft-tissue sarcoma: a multi-center study. Front Oncol 2022;12:897676. [PMID: 35814362; DOI: 10.3389/fonc.2022.897676]

[62] Jiang M, Li CL, Luo XM, Chuan ZR, Lv WZ, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. Eur J Cancer 2021;147:95-105. [PMID: 33693932; DOI: 10.1016/j.ejca.2021.01.028]

[63] Sun RJ, Fang MJ, Tang L, Li XT, Lu QY, et al. CT-based deep learning radiomics analysis for evaluation of serosa invasion in advanced gastric cancer. Eur J Radiol 2020;132:109277. [PMID: 32980726; DOI: 10.1016/j.ejrad.2020.109277]

[64] Chen H, Lin F, Zhang J, Lv X, Zhou J, et al. Deep learning radiomics to predict PTEN mutation status from magnetic resonance imaging in patients with glioma. Front Oncol 2021;11:734433. [PMID: 34671557; DOI: 10.3389/fonc.2021.734433]

[65] Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med 2018;15(11):e1002711. [PMID: 30590819; DOI: 10.1371/journal.pmed.1002711]

[66] Han W, Wang L, Lee EH, Zheng J, Zhang W, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. Eur J Nucl Med Mol Imaging 2021;48:1478-86. [PMID: 33094432; DOI: 10.1007/s00259-020-05075-4]

[67] Wang Y, Shao J, Wang P, Chen L, Ying M, et al. Deep learning radiomics to predict regional lymph node staging for hilar cholangiocarcinoma. Front Oncol 2021;11:721460. [PMID: 34765542; DOI: 10.3389/fonc.2021.721460]
[68] Lombardo E, Kurz C, Marschner S, Avanzo M, Gagliardi V, et al. Distant metastasis time to event analysis with CNNs in independent head and neck cancer cohorts. Sci Rep 2021;11:6418. [PMID: 33742070; DOI: 10.1038/s41598-021-85671-y]

[69] Kann BH, Aneja S, Loganadane GV, Kelly JR, Smith SM, et al. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. Sci Rep 2018;8:14036. [DOI: 10.1038/s41598-018-32441-y]

[70] Kim DW, Lee S, Kwon S, Nam W, Cha IH, et al. Deep learning-based survival prediction of oral cancer patients. Sci Rep 2019;9:6994. [PMID: 31061433; DOI: 10.1038/s41598-019-43372-7]

[71] Zhou P, Han J, Cheng G, Zhang B. Learning compact and discriminative stacked autoencoder for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 2019;57:4823-33. [DOI: 10.1109/TGRS.2019.2893180]

[72] Pan X, Fan YX, Yan J, Shen HB. IPMiner: hidden ncRNA-protein interaction sequential pattern mining with stacked autoencoder for accurate computational prediction. BMC Genomics 2016;17:582. [DOI: 10.1186/s12864-016-2931-8]

[73] Chang H, Han J, Zhong C, Snijders AM, Mao JH. Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications. IEEE Trans Pattern Anal Mach Intell 2018;40:1182-94. [PMID: 28129148; DOI: 10.1109/TPAMI.2017.2656884]

[74] Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, et al. Machine and deep learning methods for radiomics. Med Phys 2020;47:e185-202. [PMID: 32418336; DOI: 10.1002/mp.13678]

[75] Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019;69:127-57. [PMID: 30720861; DOI: 10.3322/caac.21552]

[76] Li S, Deng YQ, Zhu ZL, Hua HL, Tao ZZ. A comprehensive review on radiomics and deep learning for nasopharyngeal carcinoma imaging. Diagnostics (Basel) 2021;11:1523. [PMID: 34573865; DOI: 10.3390/diagnostics11091523]

[77] Tomiyama N, Johkoh T, Mihara N, Honda O, Kozuka T, et al. Using the World Health Organization classification of thymic epithelial neoplasms to describe CT findings. AJR Am J Roentgenol 2002;179:881-6. [PMID: 12239030; DOI: 10.2214/ajr.179.4.1790881]

[78] Chee CG, Yoon MA, Kim KW, Ko Y, Ham SJ, et al. Combined radiomics-clinical model to predict malignancy of vertebral compression fractures on CT. Eur Radiol 2021;31:6825-34. [PMID: 33742227; DOI: 10.1007/s00330-021-07832-x]

[79] Frood R, Burton C, Tsoumpas C, Frangi AF, Gleeson F, et al. Baseline PET/CT imaging parameters for prediction of treatment outcome in Hodgkin and diffuse large B cell lymphoma: a systematic review. Eur J Nucl Med Mol Imaging 2021;48:3198-220. [PMID: 33604689; DOI: 10.1007/s00330-021-05233-2]

[80] van Griethuysen JIM, Fedorov A, Parmar C, Hosny A, Aucoin N, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104-7. [PMID: 29092951; DOI: 10.1158/0008-5472.CAN-17-0339]

[81] Mukherjee P, Cintra M, Huang C, Zhou M, Zhu S, et al. CT-based radiomic signatures for predicting histopathologic features in head and neck squamous cell carcinoma. Radiol Imaging Cancer 2020;2:e190039. [PMID: 32550599; DOI: 10.1148/rycan.2020190039]