ORIGINAL ARTICLE

Open Access

Automated detection of lung cancer-caused metastasis by classifying scintigraphic images using convolutional neural network with residual connection and hybrid attention mechanism

Yanru Guo[1,2], Qiang Lin[1,2,3]*, Shaofang Zhao[1,2], Tongtong Li[1,2], Yongchun Cao[1,2,3], Zhengxing Man[1,2,3] and Xianwu Zeng[4]*

Abstract

Background: Whole-body bone scan is the widely used tool for surveying bone metastases caused by various primary solid tumors including lung cancer. Scintigraphic images are characterized by low specificity, bringing a significant challenge to manual analysis of images by nuclear medicine physicians. Convolutional neural network can be used to develop automated classification of images by automatically extracting hierarchal features and classifying high-level features into classes.

Results: Using convolutional neural network, a multi-class classification model has been developed to detect skeletal metastasis caused by lung cancer using clinical whole-body scintigraphic images. The proposed method consisted of image aggregation, hierarchal feature extraction, and high-level feature classification. Experimental evaluations on a set of clinical scintigraphic images have shown that the proposed multi-class classification network is workable for automated detection of lung cancer-caused metastasis, with achieving average scores of 0.7782, 0.7799, 0.7823, 0.7764, and 0.8364 for accuracy, precision, recall, F-1 score, and AUC value, respectively.

Conclusions: The proposed multi-class classification model can not only predict whether an image contains lung cancer-caused metastasis, but also differentiate between subclasses of lung cancer (i.e., adenocarcinoma and non-adenocarcinoma). On the context of two-class (i.e., the metastatic and non-metastatic) classification, the proposed model obtained a higher score of 0.8310 for accuracy metric.

Keywords: Bone scan, Skeletal metastasis, Lung cancer, Image classification, Convolutional neural network

Key points

- Automated detection of lung cancer-caused skeletal metastasis is first studied.
- Convolutional neural network is exploited to develop automated classification method.
- Clinical scintigraphic images are used to experimentally evaluate the proposed classification model.

*Correspondence: qiang.lin2010@hotmail.com; 1021619285@qq.com
[1] School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, Gansu, China
[4] Department of Nuclear Medicine, Gansu Provincial Tumor Hospital, Lanzhou, Gansu, China
Full list of author information is available at the end of the article

Guo *et al. Insights into Imaging* (2022) 13:24

Page 2 of 13

Background

Skeletal metastasis is common in several of prevalent cancers including prostate, breast, and lung cancers [1], with 80% of all skeletal metastatic lesions originating from one of these primary sites [2]. The percentage of metastasis-related death reaches up to 90% for all lung cancer mortality [3]. Early detection of skeletal metastasis is extremely important not only for decreasing morbidity but also for disease staging, outcome prediction, and treatment planning [4].

Skeletal scintigraphy (bone scan) and positron emission tomography (PET) are commonly used for surveying bone metastasis [5, 6]. Compared to PET, bone scan is more affordable and available due to its low-cost equipment and radiopharmaceutical. Bone scan is typically characterized by high sensitivity but low specificity, bringing significant challenge to manual analysis of bone scan images by nuclear medicine physicians. The reasons of low specificity are multi-fold, mainly including low spatial resolution, accumulation of radiopharmaceutical in normal skeletal structures, soft tissues or viscera, and uptake in benign processes [7].

Automated analysis of bone scan images becomes therefore desired for accurate diagnosis of skeletal metastasis. There has been a substantial amount of works aimed at developing automated diagnosis approaches using conventional machine learning models to classify bone scan images into classes [5, 8–11], where the image features were manually extracted by researchers. The handcrafted features, however, often suffer from insufficient capability and unsatisfied performance for clinical tasks [6].

Convolutional neural network (CNN), a mainstream branch of deep learning techniques, has gained huge success in automated analysis of natural images [12–14] and medical images [14–17] due to their ability to automatically extracting hierarchical features from images in an optimal way. CNN-based automated classification methods have been proposed to detect metastasis caused by a variety of various primary tumors including prostate cancer [18–23], breast cancer [22–24], lung cancer [25, 26], and both of them [25–27]. The main purpose of existing works is to develop two-class classification models to determine whether or not an image contains metastasized lesion(s) by classifying this image (normal and metastatic). Differently, a series of CNN-based methods has been proposed to classify whole-body scintigraphic images for automated detection of skeletal metastases in our previous works [28, 29], in which we did not distinguish between the primary cancers.

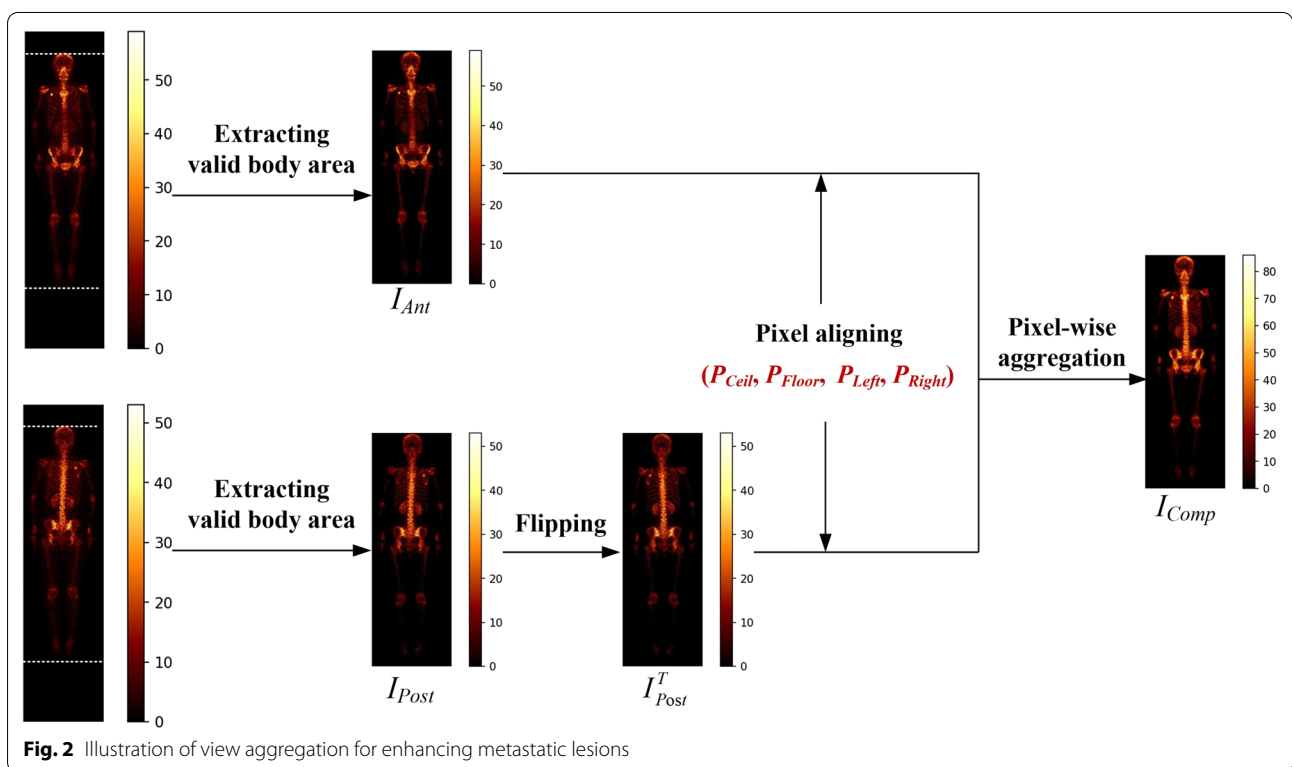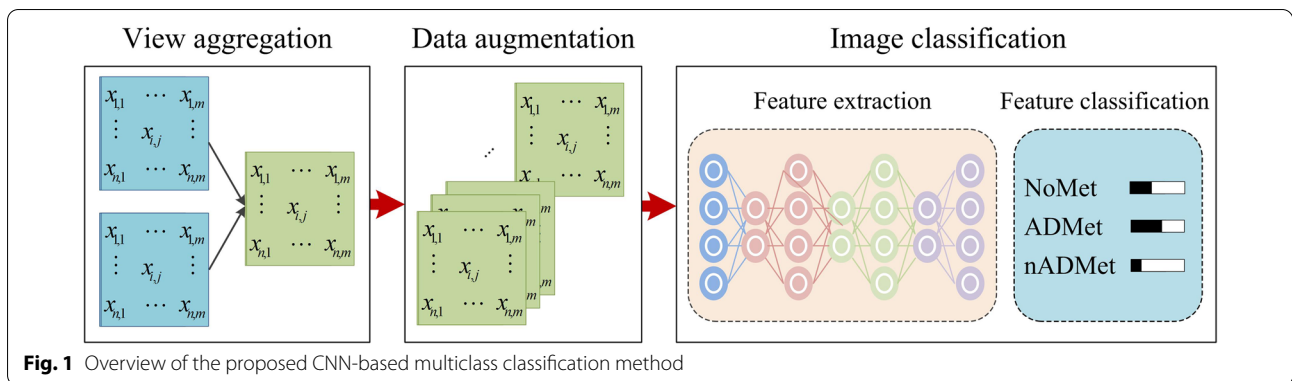Targeting at automated detection of skeletal metastasis caused by lung cancer, in this work, we propose a CNN-based multiclass classification network to classify whole-body scintigraphic images acquired from patients with clinically diagnosed lung cancer using a SPECT (single photon emission computed tomography) imaging device (i.e., GE SPECT Millennium MPR). The proposed network can not only determine whether an image contains lung cancer-caused skeletal metastasis, but also differentiate between subclasses of lung cancer (i.e., adenocarcinoma and non-adenocarcinoma).

The main contributions of this work can be summarized as: First, to the best of our knowledge, we are the first to attempt to solve the problem of automated detection of skeletal metastasis originated from various subclasses of lung cancer. Second, we convert the detection problem into the multiclass classification of low-resolution, large-size scintigraphic images using a CNN-based end-to-end network that first extracts hierarchal features from images, then aggregates these features, and finally classifies those high-level features into classes. Lastly, we use a group of scintigraphic images acquired from patients with clinically diagnosed lung cancer to evaluate the proposed method. Experimental results have shown that our CNN-based classification network performs well for distinguishing SPECT images between non-metastatic and metastatic as well as their sub-classes of metastasis.

The rest of this paper is organized as follows. We present in "Methods" section the proposed method. We report in "Results" section the experimental evaluation conducted on clinical SPECT images. In "Discussion" secton, we provide a brief discussion on the reasons that cause the misclassifications. In "Conclusions" section, we conclude this work and point out the future research directions.

Methods

To automatically detect metastasis of lung cancer in scintigraphic images, image fusion operation is first employed to enhance the lesion(s) in low-resolution whole-body scintigraphic images through aggregating the anterior- and posterior-view images of each bone scan. Parametric variation-based data augmentation is then applied to expand the size of the dataset used in this work to improve the performance of CNN-based network on classifying images as much as possible. A CNN-based end-to-end network is developed to classify the fused images by first extracting hierarchal features from images, then aggregating features, and finally classifying high-level features into classes of concerns, i.e., without metastasis (NoMet), adenocarcinoma metastasis (ADMet) and non-adenocarcinoma metastasis (nAD-Met). Figure 1 provides an overview of the proposed multiclass classification method, comprising of three main steps, i.e., view aggregation, data augmentation, and image classification.

Guo *et al. Insights into Imaging* (2022) 13:24

Page 3 of 13

Fig. 1 Overview of the proposed CNN-based multiclass classification method

Fig. 2 Illustration of view aggregation for enhancing metastatic lesions

View aggregation

During the SPECT imaging, two whole-body images were collected for each patient, corresponding to the anterior- and posteriorviews, respectively. When a primary tumor (e.g., lung cancer) invades into bone tissue, there will be an area of increased radionuclide's uptake in the image. It is common, however, the metastatic areas have varied intensity of uptake in anterior- and posterior-view images. How to enhance the metastatic areas in images becomes crucial for accurate detection of metastasis. A pixel-wise view aggregation method is proposed to 'excite' those metastatic pixels, while 'squeeze' the normal pixels by fusing two views as shown in Fig. 2.
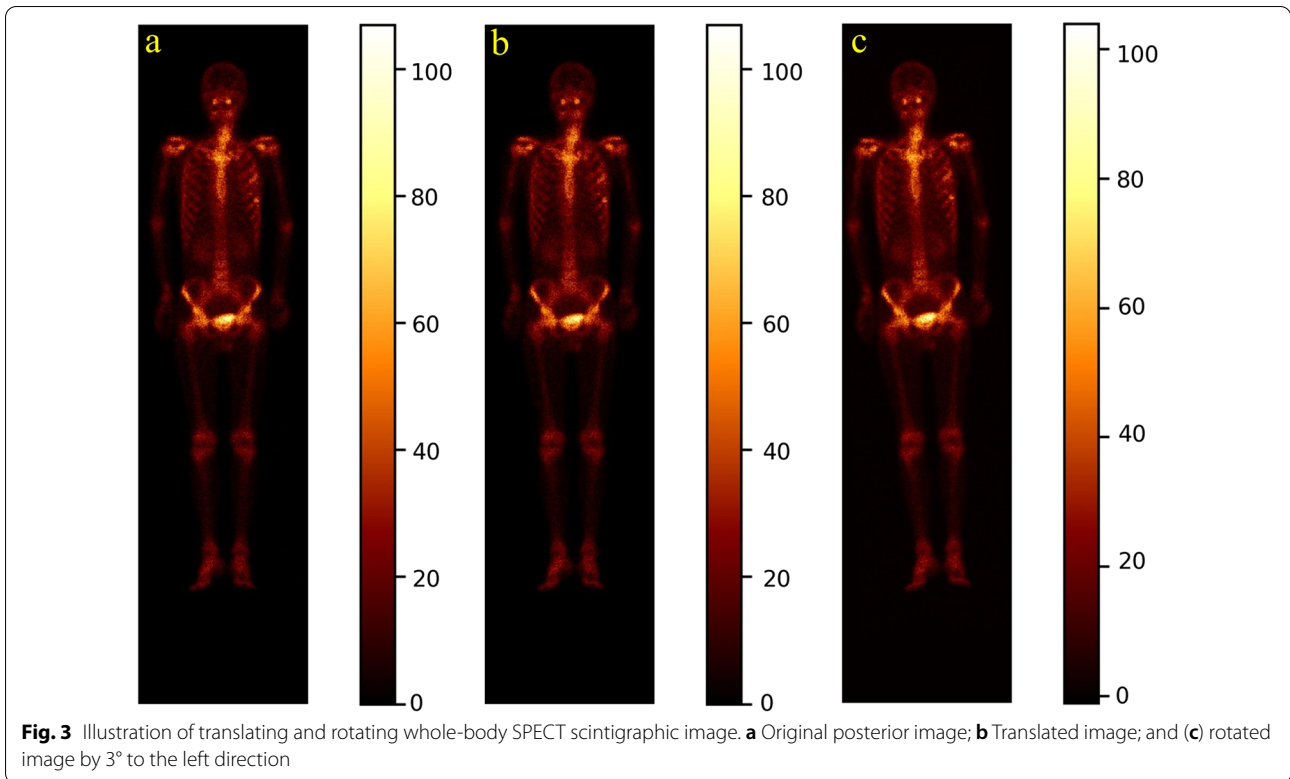
Let I_{Ant} and I_{Post} denote the anterior- and posterior-view image respectively, the pixel-wise view aggregation method works as follows.

Image flipping

The posterior-view image I_{Post} is flipped horizontally around its central vertical line to obtain an image I_{Post}^T.

Pixel aligning

A horizontal line sweeps the image (i.e., I_{Ant} and I_{Post}^T) line by line to find out the critical points P_{Ceil} and P_{Floor} by examining the pixel value that represents uptake intensity. Similarly, we use a vertical line to sweep image line

Guo *et al. Insights into Imaging*   (2022) 13:24

Page 4 of 13

Fig. 3 Illustration of translating and rotating whole-body SPECT scintigraphic image. **a** Original posterior image; **b** Translated image; and (**c**) rotated image by 3° to the left direction

by line to find out the critical points P_{Left} and P_{Right}. Two images I_{Ant} and I_{Post}^{T} will be aligned according to these four critical points.

Pixel-wise image addition

The aligned images, I_{Ant} and I_{Post}^{T}, will be aggregated to generate a composite image, I_{Comp}, according to Eq. 1.

$$I_{\text{Comp}} = f\left(I_{\text{Ant}}, I_{\text{Post}}^{T}\right) \tag{1}$$

where f is aggregation function, i.e., pixel-wise addition operation.

Data augmentation

It is widely accepted that the classification performance of CNN-based models depends on the size of dataset, with high classification accuracy always corresponding to the large dataset. Currently, a variety of various methods can be utilized to augment dataset including the parametric variation and adversarial learning techniques. In this work, we use the parametric variation technique to augment our dataset since the parametric variation-based data augmentation can obtain samples that have the same distribution as the original ones with the lower time complexity. Specifically, image translation and rotation are used, which are detailed as follows [30].

Image rotation

Given a constant $r \in [0, r_{T}]$, an image will be randomly rotated by r^{o} in either the left or right direction around its geometric center. The parameter r_{T} is experimentally determined according to the distribution of the radiotracer uptake of all images in the dataset. Figure 3d depicts the obtained image by rotating the image in Fig. 3a to the right direction by 3°.

Image translation

Given a constant $r \in [0, t_{T}]$, an image will be randomly translated by $+t$ or $-t$ pixels in either the horizontal or vertical direction. The parameter t_{T} is experimentally determined according to the distribution of the radiotracer uptake of all images in the dataset. Figure 3c shows a resulting example by translating the given image in Fig. 3a + 3 pixels horizontally.

CNN-based classification network

Table 1 outlines the structure of the proposed 26-layer CNN-based classification network, consisting of one convolution layer (Conv), one normalization layer (Norm), one pooling layer (Pool), a set of residual convolution layer attached attention (RA-Conv) with varied kernel, one global average pooling layer (GAP), and 1 Softmax layer.

Guo *et al. Insights into Imaging*   (2022) 13:24

Page 5 of 13

Table 1 Network structure of the proposed CNN-based classification model

Layer	Configuration
Conv	7×7, 64, Stride $= 2$
Norm	Batch normalization
Pool	3×3 Max pooling, Stride $= 2$
RA-Conv_2	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 2$
RA-Conv_3	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 3$
RA-Conv_5	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 5$
RA-Conv_2	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 2$
Global average pooling (GAP)	
Softmax	

An input 256×1024 scintigraphic image is convolved by the Conv layer with filter of 7×7 to calculate a feature map made of neurons, followed by a batch normalization layer and a max pooling layer with kernel of 3×3. The subsequent convolutional layers are organized as residual convolution with hybrid attention inside or outside of the convolution. A global average pooling layer is used to alleviate the over-fitting problem while speeding up the training process. The Softmax layer points out the class of an image with a real number. The main layers will be detailed as follows.
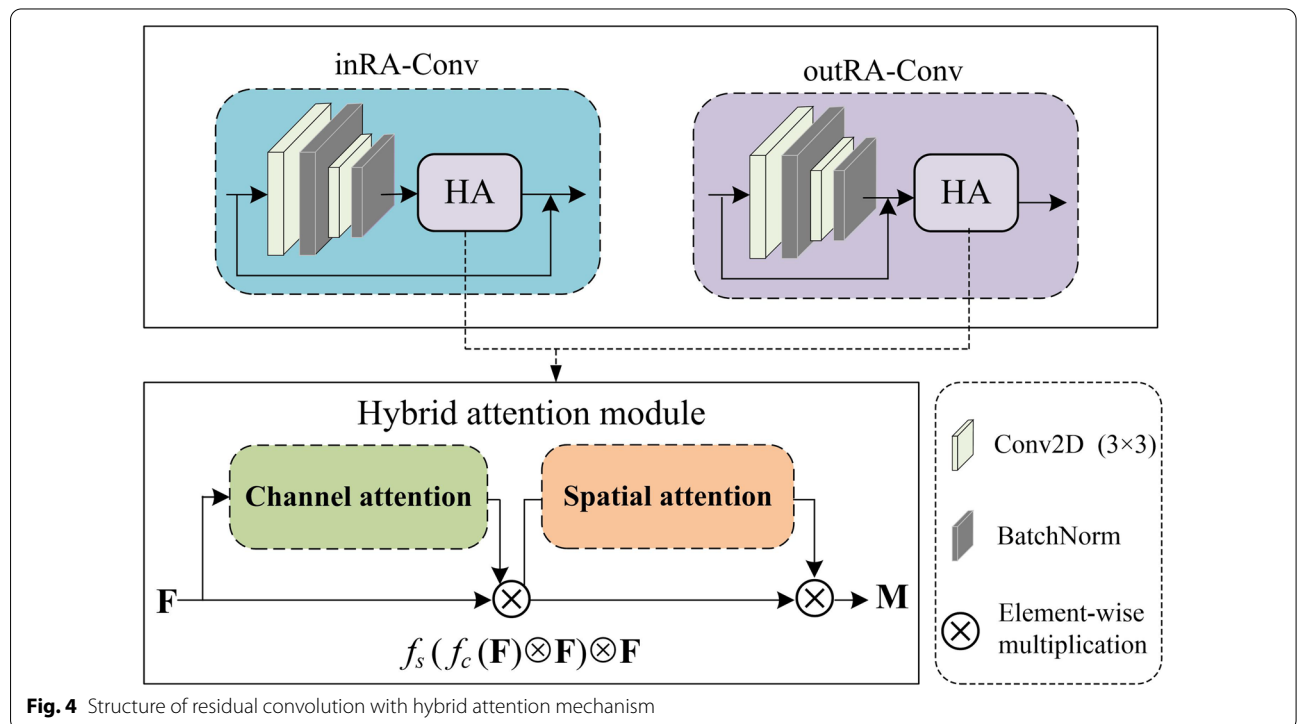
Normalization layer
Batch normalization [31] is used to accelerate the network training by making normalization a part of the model architecture and performing the normalization for each training mini-batch. With batch normalization, we can thus use much higher learning rates and be less careful about initialization.

RA-Conv layer
Figure 4 demonstrates the structure of residual convolution with hybrid attention mechanism. We use residual connection to reduce the training parameters and training time. We also introduce hybrid attention mechanism to improve network focusing on those more important areas (i.e., lesions) on the feature maps by considering only the important information. Specifically, we use inRA-Conv (outRA-Conv) to indicate that a hybrid attention module is located inside (outside) the residual convolution. The classifiers are accordingly named as Classifer-inRAC and Classifer-outRAC, respectively.

The cascaded hybrid attention module in Fig. 4 using channel and spatial attention mechanism is capable of computing complementary attention by focusing on 'what' (channel attention) and 'where' (spatial attention), respectively [32]. Specifically, let **F** be the input of a 2D

Fig. 4 Structure of residual convolution with hybrid attention mechanism

Guo *et al. Insights into Imaging* (2022) 13:24

Page 6 of 13

feature map to the channel attention sub-module. We can obtain a 1D output F, which will be further processed by the spatial attention sub-module to output a refined 2D feature map **M** according to Eq. 2.

$$\mathbf{M} = f_S\big(f_C(\mathbf{F}) \otimes \mathbf{F}\big) \otimes \mathbf{F}, \tag{2}$$

where $\otimes$ is the element-wise multiplication, and f_C and f_S denotes the channel and spatial function, respectively, which are given in Eqs. 3 and 4.

$$f_C(\mathbf{F}) = \sigma\big(MLP\big(AvgPool(\mathbf{F})\big) + MLP(MaxPool(\mathbf{F}))\big), \tag{3}$$

$$f_S(\text{F}) = \sigma\Big(f^{k \times k}\big([AvgPool(\text{F}); MaxPool(\text{F})]\big)\Big), \tag{4}$$

where σ is the sigmoid function, *MLP* is the multi-layer perceptron, *AvgPool* (*MaxPool*) is the average (max) pooling, and $f^{k \times k}$ is a convolutional operation with the kernel size of $k \times k$.

Softmax layer
The network output nodes apply the Softmax function for the number of the unordered classes. A Softmax function is defined in Eq. 5 [33].

$$f(x_j) = \frac{e^{x_j}}{\sum_{i=1}^{n} e^{x_i}}, \tag{5}$$

where $f(x_j)$ is the score of the *j*-th output node, x_j is the network input to *j*-th output node, and *n* is the number of output nodes. In fact, all of the output values $f(x)$ are a probability between 0 and 1, and their sum is 1.

Results
In this section, we provide an experimental evaluation of the proposed network using a set of clinical whole-body scintigraphic images.

Dataset
In this retrospective study, the whole-body scintigraphic images were collected from the Department of Nuclear Medicine, Gansu Provincial Tumor Hospital from Jan 2014 to Dec 2019 using a single-head gamma camera (GE SPECT Millennium MPR). SPECT imaging was performed between 2 and 3 h after intravenous injection of ^{99m}Tc-MDP (20–25 mCi) using a parallel-beam low-energy high-resolution (LEHR) collimator (energy peak$=140$ keV, intrinsic energy resolution $\leq 9.5\%$, energy window$=20\%$, and intrinsic spatial resolution ≤ 6.9 mm). Each SPECT image was stored in a DICOM (Digital Imaging and Communications in Medicine) file with the imaging size of 256×1024. Every element in an image is represented by a 16-bit unsigned integer, differing from the natural images in which element ranges from 0 to 255.

A total of 506 patients who were clinically diagnosed with lung cancer were encompassed in this study. Figure 5 demonstrates the distribution of patients with respect to gender and age.

There are 1011 images collected from 506 patients due to the phenomenon of images being not successfully recorded. We categorize all these images into three subclasses, i.e., NoMet (n$=614$, $\approx 60.73\%$), ADMet (n$=237$, $\approx 23.44\%$), and nADMet (n$=160$, $\approx 15.83\%$).

To keep the balance between samples in different subclasses, we randomly selected 226 images from NoMet class and group the original images into dataset D1 as shown in Table 2. Applying data augmentation technique on D1, we obtained an augmented dataset D2. The dataset D3 is achieved by aggregating images in D2.

For supervised image classification problem, the CNN-based model is evaluated by comparing the automated classification results against ground truth (human performance) that is often obtained by manually labeling images. However, it is time-consuming, laborious, and subjective to manually label low-resolution, large-size SPECT images. To facilitate labeling SPECT image, in this work, we developed an annotation system based on the LabelMe (http://labelme.csail.mit.edu/Release3.0/) released by MIT.

With LabelMe-based annotation system, imaging findings including the DICOM file and the textual diagnostic report can be imported into the system in advance. In the labeling process, three nuclear medicine physicians from the Department of Nuclear Medicine, Gansu Provincial Tumor Hospital manually labeled areas on the visual presentation of DICOM file with a shape tool (e.g., polygon and rectangle). The labeled area will be annotated with a self-defined code combined with the name of disease or body part. The manually labeled results for all images act as ground truth in the experiments and form an annotation file together to feed into the classifiers.

Experimental setup
The evaluation metrics we use are *accuracy*, *precision*, *recall*, *specificity*, *F*-1 score, and AUC (Area Under ROC Curve), which are defined in Eqs. 6–10.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$
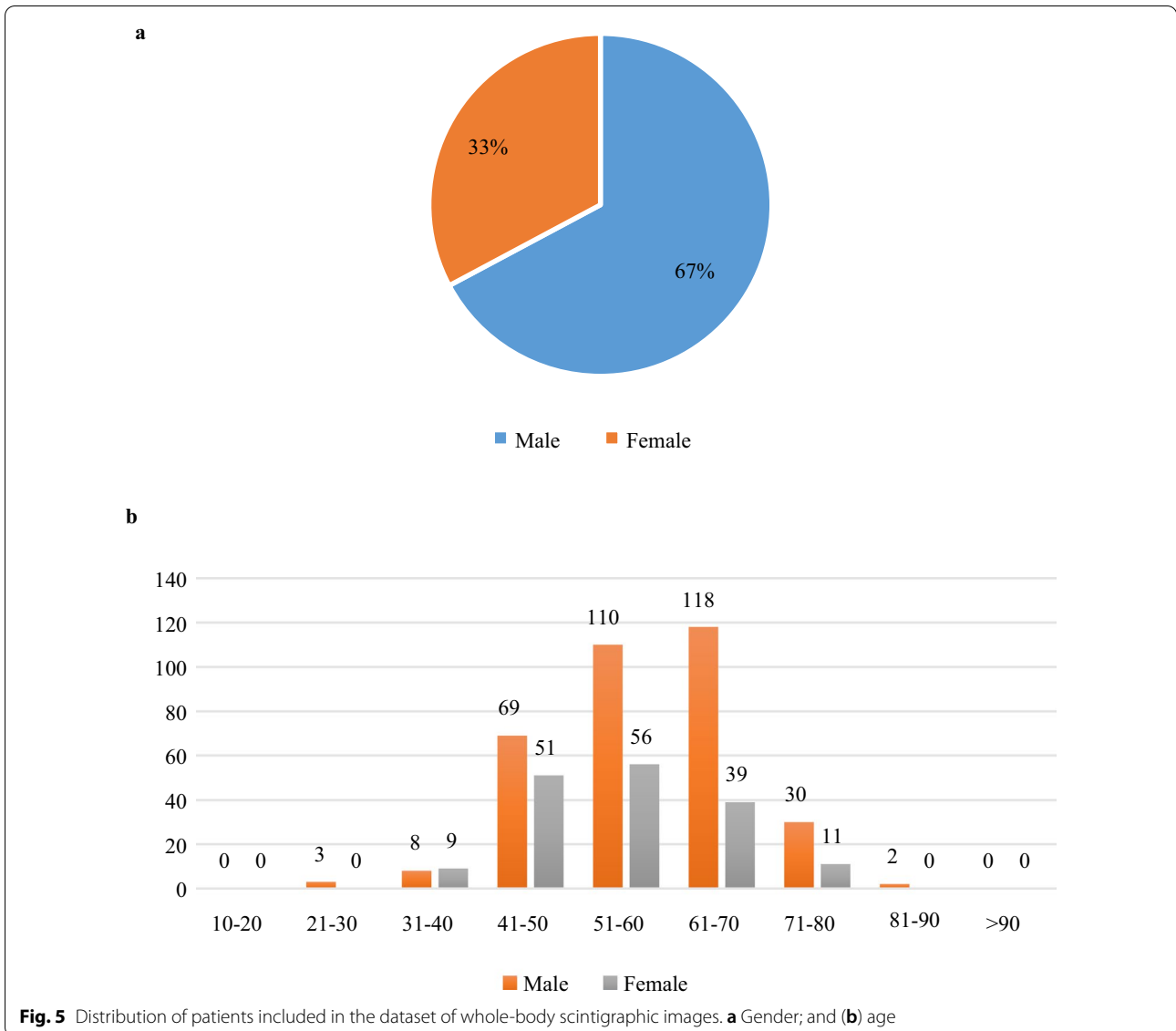
$$\text{Precision} = \frac{TP}{TP + FP}, \tag{7}$$

Guo *et al. Insights into Imaging* (2022) 13:24

Page 7 of 13

Fig. 5 Distribution of patients included in the dataset of whole-body scintigraphic images. **a** Gender; and (**b**) age

Table 2 An overview of the datasets used in this work

Dataset	ADMet	nADMet	NoMet	Total
D1	237	160	226	623
D2	624	640	614	1878
D3	318	320	307	945

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{8}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{9}$$

$$F - 1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{10}$$

where the notations are $TP =$ True Positive, $TN =$ True Negative, $FP =$ False Positive and $FN =$ False Negative.

It is desirable that a classifier shows both a high true positive rate ($TPR = Recall$), and a low false positive rate ($FPR = 1 – Specificity$) simultaneously. The ROC curve shows the true positive rate (y-axis) against the false positive rate (x-axis), and the AUC value is defined as the area under the ROC curve. As a statistical explanation, the AUC value is equal to the probability that a randomly chosen positive image is ranked higher than a randomly chosen negative image. Therefore, the closer to 1 the AUC value is, the higher performance the classifier achieves.

Guo *et al. Insights into Imaging*   (2022) 13:24

Page 8 of 13

Table 3 Parameters setting of the proposed classification network

Parameter	Value
Learning rate	0.01
Optimizer	Adam
Batch size	32
Epoch	300

Table 4 Scores of evaluation metrics obtained by Classifer-inRAC and Classifer-outRAC on testing samples in dataset D3

Classifier	Accuracy	Precision	Recall	F-1 score
Classifer-inRAC	**0.7782**	**0.7799**	**0.7823**	**0.7764**
Classifer-outRAC	0.6725	0.7233	0.6831	0.6723

Best value in each column is highlighted in bold

Table 5 Scores of evaluation metrics obtained by Classifer-inRAC on the testing samples in datasets D1, D2, and D3

Dataset	Accuracy	Precision	Recall	F-1 score
D1	0.6150	0.6324	0.6227	0.6058
D2	0.6968	0.7001	0.7024	0.6930
D3	**0.7782**	**0.7799**	**0.7823**	**0.7764**

Best value in each column is highlighted in bold

We divided every dataset (D1, D2 and D3) into two parts: training set and testing set, with the ratio of them being 7: 3. It means that we use 70% of samples in each dataset to train the classifiers, and the rest 30% for testing the classifiers. Images including the augmented ones from the same patient were not divided into the different subsets because they would show similarities. The parameters setting is shown in Table 3.

The experiments are run in Tensorflow 2.0 on an Intel Core i7-9700 PC with 32 GB RAM running Windows 10.

Experimental results

For the proposed multiclass classifiers Classifer-inRAC and Classifer-outRAC, Table 4 reports the scores of the defined evaluation metrics obtained on the testing samples in dataset D3.

Table 4 shows that the classifier Classifer-inRAC performs better than Classifer-outRAC. Results in Table 5 further show that Classifer-inRAC obtains the best performance on the aggregated samples in augmented dataset (i.e., D3).
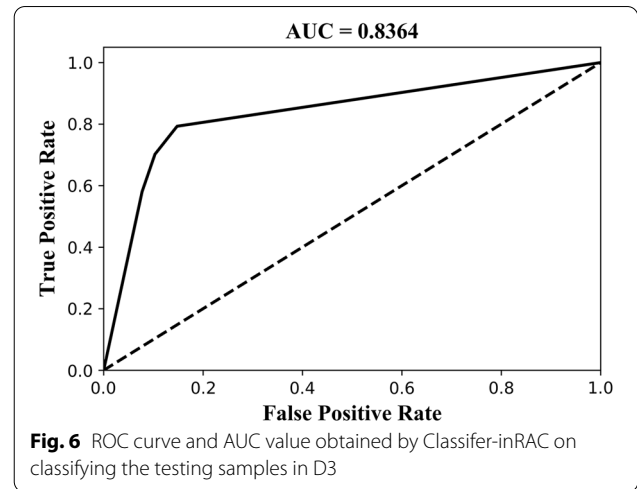
Fig. 6 ROC curve and AUC value obtained by Classifer-inRAC on classifying the testing samples in D3
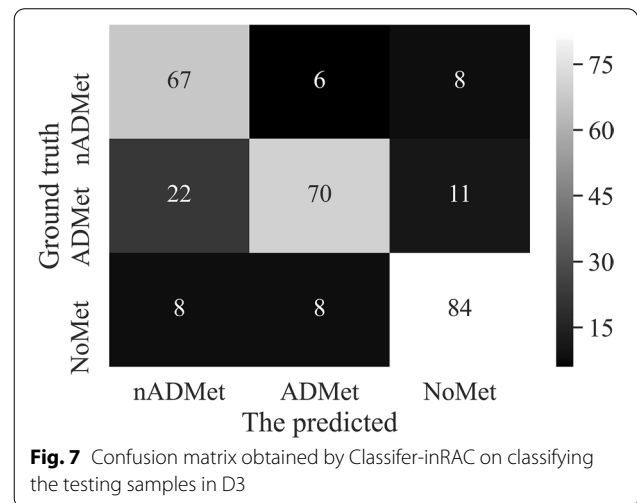
Fig. 7 Confusion matrix obtained by Classifer-inRAC on classifying the testing samples in D3
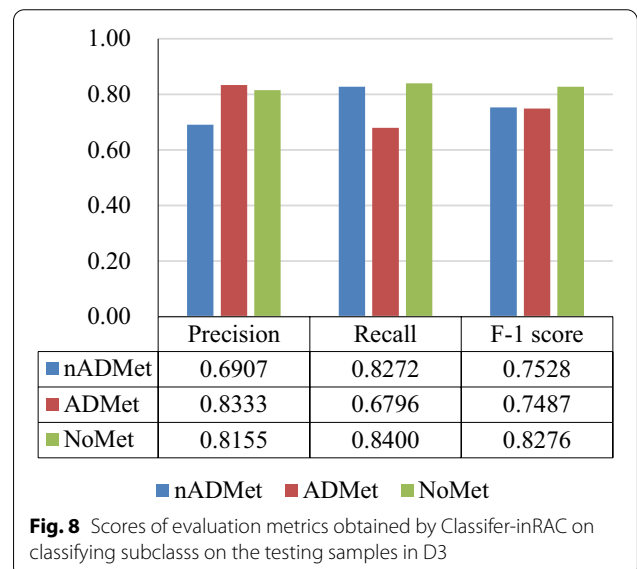

	Precision	Recall	F-1 score
nADMet	0.6907	0.8272	0.7528
ADMet	0.8333	0.6796	0.7487
NoMet	0.8155	0.8400	0.8276

Fig. 8 Scores of evaluation metrics obtained by Classifer-inRAC on classifying subclasss on the testing samples in D3

Guo *et al. Insights into Imaging* (2022) 13:24

Page 9 of 13

Table 6 Effects of network structure on classification performance obtained on dataset D3

Residual	Attention	Accuracy	Precision	Recall	F-1 score
×	×	0.6937	0.7032	0.7000	0.6940
×	√	0.7042	0.7416	0.7047	0.7031
√	×	0.7500	0.7614	0.7532	0.7497
√	√	**0.7782**	**0.7799**	**0.7823**	**0.7764**

Best value in each column is highlighted in bold

Table 7 Overview of classifiers with similar structure but different depth from Classifer-inRAC

Layer	Clasifier-18 Configuration	Clasifier-34	Clasifier-inRAC
Conv	7×7, 64, Stride $= 2$		
Norm	Batch normalization		
Pool	3×3 Max pooling, Stride $= 2$		
RA-Conv	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 2$
RA-Conv	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 3$
RA-Conv	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 5$
RA-Conv	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 2$

Global average pooling (GAP)

Softmax

Fig. 9 Classfication performance comparison between different classifiers in Table 7

Table 8 Two-class classification performance obtained by Classifer-inRAC

Accuracy	Precision	Recall	F-1 score	AUC
0.8310	0.8696	0.8696	0.8696	0.8147
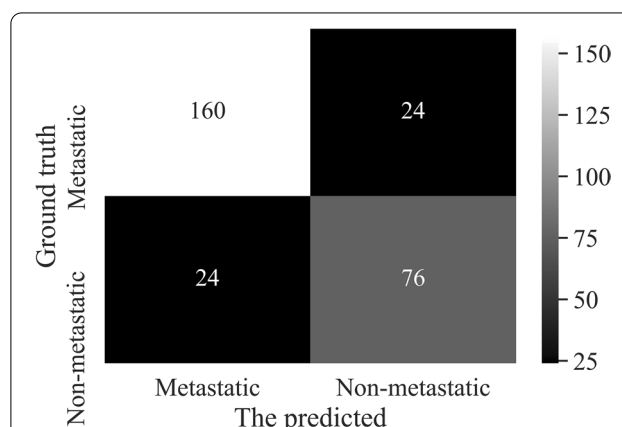
Fig. 10 Confusion matrix of two-class classification obtained by Classifer-inRAC

Figure 6 shows the ROC curve and AUC value obtained by Classifer-inRAC on classifying the testing samples in D3, where AUC value = 0.8364.

We further examine the ability of Classifer-inRAC on differentiating between the subclasses of images in dataset D3 by providing confusion matrix in Fig. 7 and scores of evaluation metrics in Fig. 8.

Experimental results in Figs. 7 and 8 show that differentiating subclasses between images with metastasis is more challenging than differentiating between metastatic and non-metastatic images. There are 22 ADMet images that have been incorrectly identified as nADMet ones.

With the testing samples in dataset D3, we show the impacts of network structure and depth on classification performance obtained by the proposed classifer Classifer-inRAC.

Table 6 reports the scores of evaluation metrics obtained by Classifer-inRAC after we remove the residual connection and hybrid attention module from Classifer-inRAC.
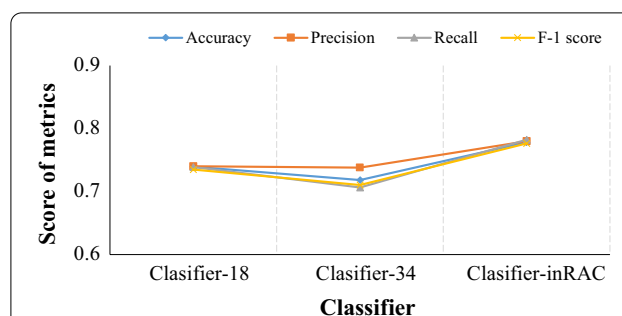
It shows that the best performance can be obtained if Classifer-inRAC has residual connection and hybrid attention module simultaneously from the scores of evaluation metrics as shown in Table 6. Separately, residual connection has more positive impact than hybrid attention mechanism on the classification performance.

Following the architectural design of Classifer-inRAC, we define two classifiers with different network depth, which are given in Table 7.

Figure 9 reports the scores of evaluation metrics obtained by the classifiers defined in Table 7 and Classifer-inRAC, showing comparative advantage of the proposed classifier on classifying whole-body images.

Guo *et al. Insights into Imaging*   (2022) 13:24

Page 10 of 13

Table 9 An overview of two classical CNNs-based models used for comparative analysis

Model	Number of weight layers	Filter	Activation	Learning rate
Inception-v1	9 Inception blocks	$1 \times 1, 3 \times 3, 5 \times 5$	ReLU	10^{-2}
VGG 11	11	3×3	ReLU	10^{-2}

Table 10 Scores of evaluation metrics obtained by the proposed model and two classical models

Model	Accuracy	Precision	Recall	F-1 score
Inception v1	0.5387	0.6003	0.5490	0.5415
VGG 11	0.7324	0.7309	0.7333	0.7309
Classifer-inRAC	**0.7782**	**0.7799**	**0.7823**	**0.7764**

Best value in each column is highlighted in bold

We further test the performance of Classifer-inRAC on two-class classification by merging the metastatic subclasses (i.e., ADMet and nADMet) in the dataset D3. Specifically, the dataset for two-class classification is consisted of metastatic images ($n = 638$, $\approx 67.51\%$) and non-metastatic images ($n = 307$, $\approx 32.49\%$). Table 8 reports the scores of evaluation metrics on two-class classification of testing samples and Fig. 10 depicts the corresponding confusion matrix.

The results of two-class classification show that our classifier performs better on differentiating between metastatic and non-metastatic images than classifying images in different subclasses.

A comparable analysis has also been performed between the proposed model and two classical deep models Inception-v1 [34] and VGG 11 [35], which are given in Table 9 by providing their network structures.

The scores of evaluation metrics obtained by three classifiers on the dataset D3 are reported in Table 10, showing that our model is more suitable for classifying lung cancer-caused images than the classical models. The possible reason is that the network structure of our model (i.e., residual convolution combined with hybrid attention) is capable of extracting more representative features of metastatic lesions.
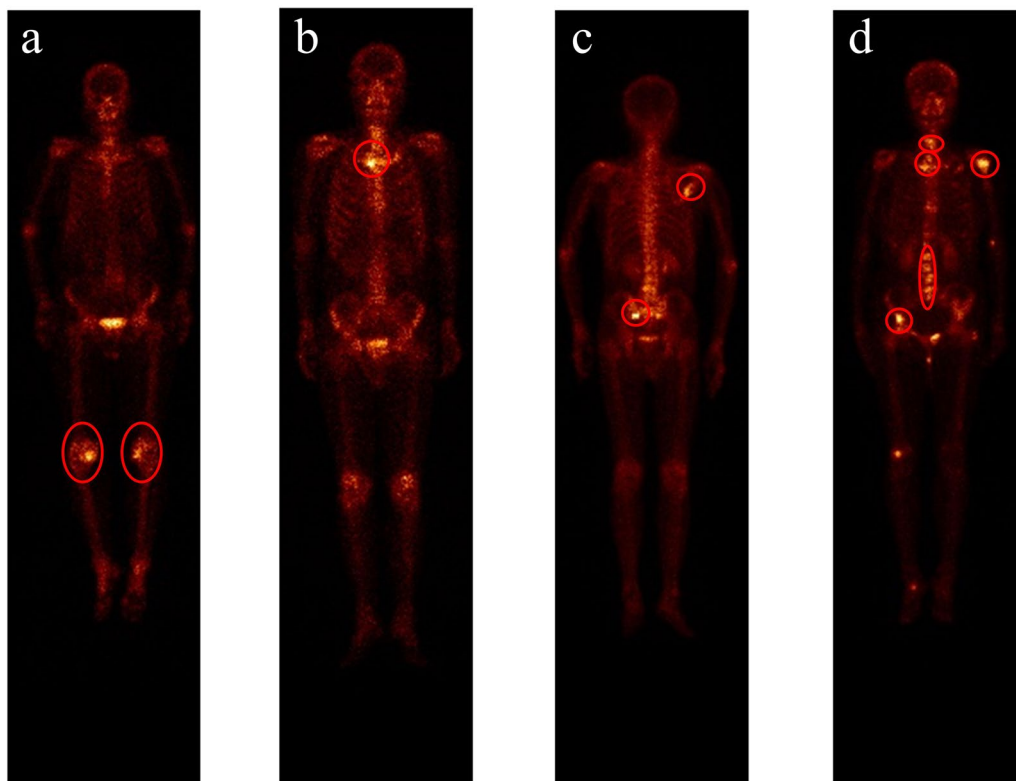
Fig. 11 An illustration of the classified images by multiclass classifier Classifer-inRAC. **a** NoMet incorrectly detected as metastatic; **b** ADMet incorrectly detected as nADmet; **c** Correctly detected nADMet image; and (**d**) Correctly detected ADMet image
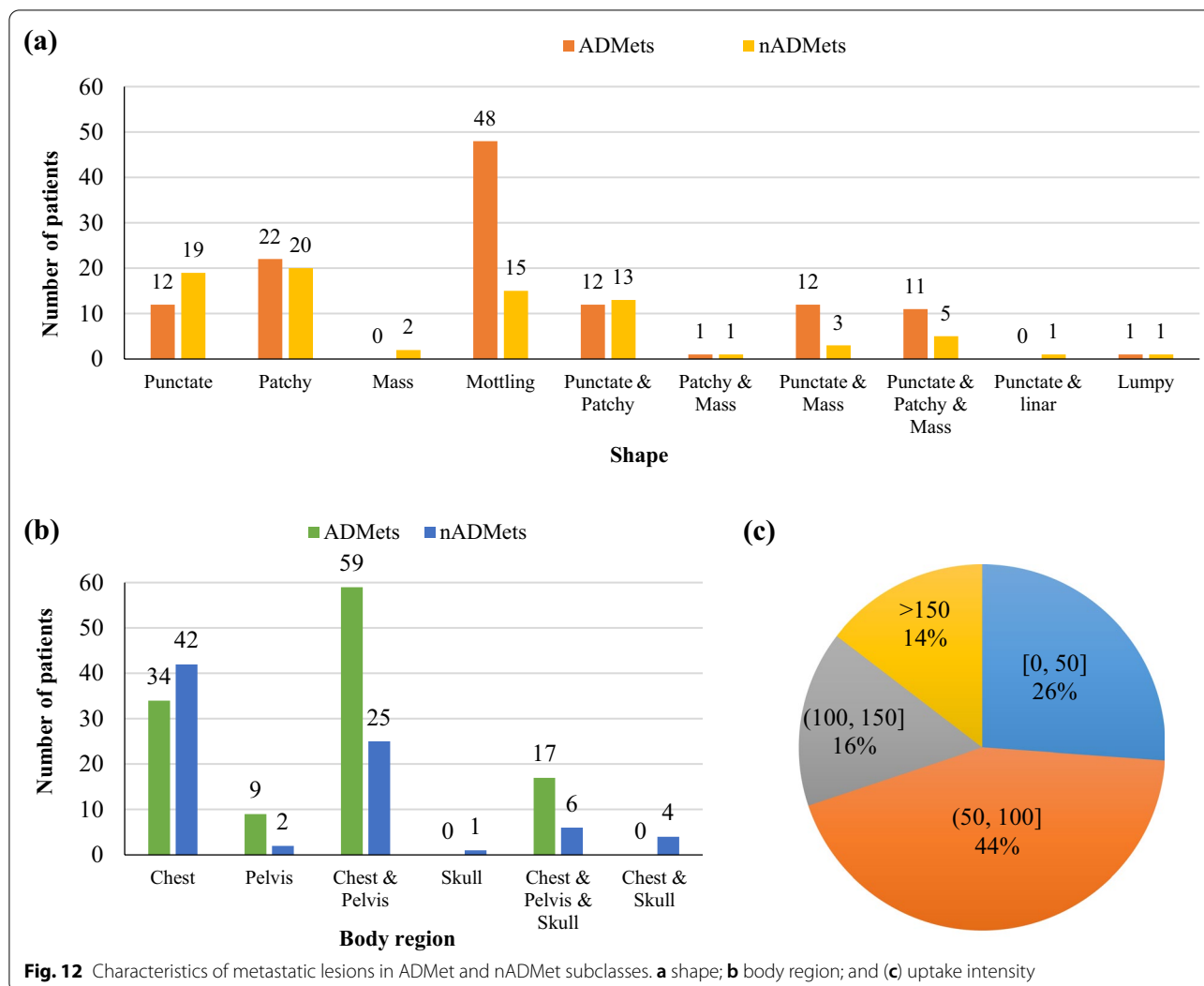
Fig. 12 Characteristics of metastatic lesions in ADMet and nADMet subclasses. **a** shape; **b** body region; and (**c**) uptake intensity

Discussion

In this section, we provide a brief discussion about the reasons that may cause the misclassifications by providing a group of examples in Fig. 11.

Now, we provide the reasons for misclassification explained by one nuclear medicine physician and one oncologist from Gansu Provincial Tumor Hospital.

Misclassification between the metastatic and non-metastatic

Uptake of ^{99m}Tc-MDP in benign processes (i.e., knee arthritis) is detected as metastatic lesions by the developed classifier due to the visually similar appearances to skeletal metastasis (see Fig. 11a). Furthermore, a normal bone would show a higher concentration of activity in trabecular bone with a large-mineralizing surface area like the spine. This brings huge challenge to the CNN-based automated classification of SPECT images,

hence the metastatic images being misclassified as non-metastatic.

Misclassification between the diseased subclasses

It is very challenging to accurately classify metastatic images since skeletal metastases are often distributed irregularly in the axial skeleton and typically show variability in size, shape, and intensity [7]. The irregularly distributed radioactivity of ADMet can mimic nADMet, and vice versa, resulting in misclassification between ADMet and nADMet (see Fig. 11b).

Multiclass classification vs. two-class classification

Multiclass classification aims to not only determine whether an image contains lung cancer-caused skeletal metastasis, but also differentiate between subclasses of lung cancer (i.e., ADMet and nADMet). This is more difficult than to answer that an image whether contains

Guo *et al. Insights into Imaging*   (2022) 13:24

Page 12 of 13

metastasis (i.e., two-class classification). So, the proposed classifier Classifer-inRAC obtained score of 0.8310 and 0.7782 for accuracy metric for multiclass and tow-class classification problems, respectively.

Metastatic lesions are further examined by providing statistical analysis of shape, location (body region), and uptake intensity in Fig. 12. The mottling, patchy, punctate lesions dominate both the ADMet and nADMet metastasis as shown in Fig. 12a. The chest (vertebra and ribs) acts the main location (i.e., body region) where the lung cancer-caused metastasis is frequently present in as shown in Fig. 2b. As shown in Fig. 12c, the distribution of detected uptake intensity ranges widely, with 44% of lesions falling into [50, 100]; and much higher uptake can often be detected in the regions of urinary bladder and injection point. This further reveals that it is more difficult to develop an automated method for analyzing scintigraphic images than natural images in which the value of pixel ranges from 0 to 255.

To alleviate the issues mentioned above, technical solutions need to be developed in the future. With a large-scale dataset of SPECT images, representative image features can be extracted for each kind of subclasses by CNN-based end-to-end classifiers. This would contribute to improving the performance of distinguishing between metastatic and non-metastatic images. Moreover, statistical analysis conducted on large-scale SPECT images and pathologic findings would have the potential to develop a multi-modal fusion classifier, enabling to achieve higher classification performance between metastatic images caused by various subclasses of lung cancer.

Conclusions

Targeting the automated detection of lung cancer-caused metastasis with SPECT scintigraphy, we have developed a convolutional neural network with the hybrid attention mechanism in this work. Parametric variation was first conducted to augment the dataset of original images. An end-to-end CNN-based classification network has been proposed to automatically extract features from images, aggregate features, and classify high-level features into classes. Clinical whole-body scintigraphic images were utilized to evaluate the developed network. Experimental results have demonstrated that our self-defined network performs well in detecting lung cancer-caused metastasis as well as differentiating between subclasses of lung cancer. The analysis has also been conducted to compare the proposed model with other related models. The results reveal that our method can be used for determining whether an image contains lung cancer-caused skeletal metastasis and differentiating between subclasses of lung cancer.

In the future, we plan to extend our work in the following directions. First, we intend to collect more data of images and laboratory findings to improve the proposed multiclass classification model. Hopefully, a robust and effective computer-aided diagnosis system will be developed. Second, we attempt to develop deep learning-based approaches that can classify whole-body SPECT images with multiple lesions from various primary diseases that may present in a single image.

Author details
[1]School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, Gansu, China. [2]Key Laboratory of Streaming Data Computing Technologies and Application, Northwest Minzu University, Lanzhou, Gansu, China. [3]Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu, China. [4]Department of Nuclear Medicine, Gansu Provincial Tumor Hospital, Lanzhou, Gansu, China.

References
1. Valkenburg KC, Steensma MR, Williams BO et al (2013) Skeletal metastasis: treatments, mouse models, and the Wnt signaling. Chin J Cancer 32(7):380–396
2. Hess KR, Varadhachary GR, Taylor SH et al (2006) Metastatic patterns in adenocarcinoma. Cancer 106:1624–1633
3. Mehlen P, Puisieux A (2006) Metastasis: a question of life or death. Nat Rev Cancer 6(6):449–458
4. Chang CY, Gill CM, Simeone FJ, et al (2014) Comparison of the diagnostic accuracy of 99m-Tc-MDP bone scintigraphy and 18F-FDG PET/CT for the detection of skeletal metastases. Acta Radiol 58:1–8
5. Sadik M, Jakobsson D, Olofsson F et al (2006) A new computer-based decision-support system for the interpretation of bone scans. Nucl Med Commun 27(5):417–423
6. Shan H, Jia X, Yan P, LiY, Paganetti H, Wang G (2020) Synergizing medical imaging and radiotherapy with deep learning. Mach Learn: Sci Technol 1(2): 021001
7. Nathan M, Gnanasegaran G, Adamson K, Fogelman I (2013) Bone scintigraphy: patterns, variants, limitations and artefacts. Springer, Berlin
8. Sadik M, Hamadeh I, Nordblom P, Suurkula M, Hoglund P, Ohlsson M, Edenbrandt L (2008) Computer–assisted interpretation of planar whole-body bone scans. J Nucl Med 49(12):1958–1965
9. Aslanta A, Dandl E, Akrolu M (2016) CADBOSS: a computer-aided diagnosis system for whole-body bone scintigraphy scans. J Cancer Res Ther 12(2):787–792
10. Mac A, Fgeb C, Svp D (2021) Object-oriented classification approach for bone metastasis mapping from whole-body bone scintigraphy. Phys Med 84:141–148
11. Elfarra FG, Calin MA, Parasca SV (2019) Computer-aided detection of bone metastasis in bone scintigraphy images using parallelepiped classification method. Ann Nucl Med 33(11):866–874
12. Algan G, Ulusoy I (2021) Image classification with deep learning in the presence of noisy labels: a survey. Knowl-Based Syst 215(3): 106771
13. Minaee S, Boykov YY, Porikli F, Plaza AJ, Terzopoulos D (2021) Image segmentation using deep learning: a survey. IEEE Trans Softw Eng 99:1–1
14. Taghanaki SA, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G (2021) Deep semantic segmentation of natural and medical images: a review. Artif Intell Rev 54:137–178
15. Lin Q, Man Z, Cao Y et al (2020) Classifying functional nuclear images with convolutional neural networks: a survey. IET Image Proc 14(14):3300–3313
16. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42(9):60–88
17. Yin XH, Wang YC, Li DY (2021) Survey of medical image segmentation technology based on U-Net structure improvement. J Softw 32(2):519–550
18. Dang J (2016) Classification in none scintigraphy images using convolutional neural networks. Lund University
19. Papandrianos N, Papageorgiou E, Anagnostis A, Papageorgiou K (2020) Bone metastasis classification using whole body images from prostate cancer patients based on convolutional neural networks application. PLoS One 15(8):e0237213
20. Papandrianos N, Papageorgiou E, Anagnostis A, Papageorgiou K (2020) Efficient none metastasis diagnosis in bone scintigraphy using a fast convolutional neural network architecture. Diagnostics 10(8):532
21. Papandrianos N, Papageorgiou E, Anagnostis A (2020) Development of convolutional neural networks to identify bone metastasis for prostate cancer patients in bone scintigraphy. Ann Nucl Med 34:824–832
22. Cheng DC, Hsieh TC, Yen KY, Kao CH (2021) Lesion-based bone metastasis detection in chest bone scintigraphy images of prostate cancer patients using pre-train, negative mining, and deep learning. Diagnostics 11(3):518
23. Cheng DC, Liu CC, Hsieh TC, Yen KY, Kao CH (2021) Bone metastasis detection in the chest and pelvis from a whole-body bone scan using deep learning and a small dataset. Electronics 10:1201
24. Papandrianos N, Papageorgiou E, Anagnostis A, Papageorgiou K (2020) A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans. Appl Sci 10(3):997
25. Pi Y, Zhao Z, Xiang Y, Li Y (2020) Automated diagnosis of bone metastasis based on multi-view bone scans using attention-augmented deep neural networks. Med Image Anal. 65: 101784
26. Zhao Z, Pi Y, Jiang L, Cai H (2020) Deep neural network based artificial intelligence assisted diagnosis of bone scintigraphy for cancer bone metastasis. Sci Rep 10:17046
27. Lin Q, Li T, Cao C, Cao Y, Man Z, Wang H (2021) Deep learning based automated diagnosis of bone metastases with SPECT thoracic bone images. Sci Rep 11:4223
28. Lin Q, Cao C, Li T, Cao Y, Man Z, Wang H (2021) Multiclass classification of whole-body scintigraphic images using a self-defined convolutional neural network with attention modules. Med Phys 48:5782–5793
29. Lin Q, Cao C, Li T, Cao Y, Man Z, Wang H (2021) dSPIC: a deep SPECT image classification network for automated multi-disease, multi-lesion diagnosis. BMC Med Imaging 21:122
30. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1):60
31. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: 32nd International conference on international conference on machine learning (ICML), Lille, France, July 6–11, 2015
32. Woo S, Park J, Lee JY, et al. (2018) CBAM: convolutional block attention module. arXiv:1807.06521v2
33. Chen B, Deng W, Du J (2017) Noisy Softmax: improving the generalization ability of DCNN via postponing the early softmax saturation. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, USA, July 21–26, 2017
34. Szegedy C, Liu W, Jia Y, et al (2015) Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, June 8–10, 2015
35. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.