



# Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET

Inês Domingues<sup>1</sup> · Gisèle Pereira<sup>2</sup> · Pedro Martins<sup>2</sup> · Hugo Duarte<sup>1</sup> · João Santos<sup>1</sup> · Pedro Henriques Abreu<sup>2</sup> 

Published online: 21 November 2019  
© Springer Nature B.V. 2019

## Abstract

Medical imaging is a rich source of invaluable information necessary for clinical judgements. However, the analysis of those exams is not a trivial assignment. In recent times, the use of deep learning (DL) techniques, supervised or unsupervised, has been empowered and it is one of the current research key areas in medical image analysis. This paper presents a survey of the use of DL architectures in computer-assisted imaging contexts, attending two different image modalities: the actively studied computed tomography and the under-studied positron emission tomography, as well as the combination of both modalities, which has been an important landmark in several decisions related to numerous diseases. In the making of this review, we analysed over 180 relevant studies, published between 2014 and 2019, that are sectioned by the purpose of the research and the imaging modality type. We conclude by addressing research issues and suggesting future directions for further improvement. To our best knowledge, there is no previous work making a review of this issue.

**Keywords** Deep learning · Computed tomography · Positron emission tomography · Medical imaging

## 1 Introduction

Medical imaging has been an integral part of medical procedures, used for early detection, diagnosis, monitoring, and treatment assessment of different medical conditions in a non-invasive manner (Kasban et al. 2015). It involves different techniques to visualise biological processes. These imaging techniques are called imaging modalities and are divided into two major paradigms—anatomical or structural modalities and functional modalities (Bushberg and Boone 2011). Each type of technology provides distinctive information about the partic-

---

✉ Pedro Henriques Abreu  
pha@dei.uc.pt

<sup>1</sup> CI-IPOP, IPO Porto, Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal

<sup>2</sup> CISUC, Department of Informatics Engineering, University of Coimbra, Pinhal de Marrocos, 3030-290 Coimbra, Portugal

ular area of the body being studied. In the spectrum of a broad range of imaging modalities are the specialities of computed tomography (CT) and positron emission tomography (PET).

Most interpretations of medical images require experienced clinicians. Given wide variations of interpretations of images leading to different diagnoses and being a laborious task which may be translated into high costs and human errors, health professionals can benefit from computer-assisted interventions (Chen et al. 2013; Li and Nishikawa 2015). Consequently, machine learning associated with image processing techniques has been widely used in the domain of medical imaging analysis by assisting image interpretation (Shen et al. 2017a).

DL has become a dominant research topic with applications in countless fields. It can be seen as an improvement of artificial neural network (ANNs) with more layers of virtual neurons increasing the predictive power from data. Data representations or features have been traditionally extracted from examples within a pre-processing step. The major advantage of DL techniques is the detection of relevant data representations in a self-taught way. Instead of extracting features manually, it can learn from examples enabling the process to map from raw input images to the final classification. There is no loss of information occurring due to indirect extraction processes that are commonly characterised by feature extraction from previously segmented objects.

There are several strategies to employ deep models. One application is the train from scratch or full training. Given an input, the architecture is trained in an end-to-end way to ultimately be used as a classifier. However, effective model training has obstacles (Tajbakhsh et al. 2016). It requires large computational and memory resources, and a huge amount of labelled data which are not always available and are difficult to obtain in the medical domain, making the dataset not adequate to train a full network on. Moreover, by training deep architectures, over-fitting and convergence issues often arise (Erhan et al. 2009). A possible direction to be explored is the extraction of image features from the input to learn representations from visual objects (Bengio et al. 2013). The challenge of this usage is to interpret the learned features and discern whether or not the learned features can be used alone for the task or if it is necessary to combine them with handcrafted features (Greenspan et al. 2016). Furthermore, a system can apply knowledge learned from prior tasks to new fields that reduce the cost of optimising parameters and the network architecture. A pre-trained network can be exploited either directly or by transfer learning (Greenspan et al. 2016). Two transfer learning strategies have been used. Fine-tuning a pre-trained network on medical data and using a pre-trained network as a feature extractor (Litjens et al. 2017).

DL has established noteworthy progress and its applications have spread to the entire field of healthcare. Medical image analysis is not an exception. Other works have studied the application of DL in general medical image analysis (Litjens et al. 2017; Shen et al. 2017a), in modalities such as CT and magnetic resonance imaging (MRI) (Jung et al. 2017; Erickson et al. 2017) and in molecular images particularly in cancer (Xue et al. 2017). CT is a popular and firmly studied modality in medical imaging, in contrast to PET. The validation of DL in PET context is an under-studied topic. These modalities have a direct impact on the final diagnosis, staging and treatment planning in many diseases and have been widely adopted in the clinical practice. It is thus of great importance that the information extracted from each modality is exploited and discussed to understand further directions to integrate both modalities. The major contribution of this paper lies in the exploitation of DL techniques in the application of advanced imaging analysis. We believe that this survey not only would be helpful for younger researchers, as it provides a comprehensive review on the application of DL for CT, PET and the use of both CT and PET, but also for senior researchers who want to explore future directions detected in the several analysed research problems. To our best

knowledge, there is no previous work making a review of this issue. We opted not to include other imaging modalities since there are already covered in the previously mentioned works.

The remainder of this paper is organised as follows. Section 2 covers the main DL techniques that have been used for medical image analysis. Section 3 analysis the distribution of the reviewed works. Section 4 depicts relevant works about contributions of DL to segmentation, classification, detection and other important tasks in medical image analysis. Finally, in Sect. 5 some conclusions and future directions are discussed.

## 2 Preliminaries

The concepts of the main DL techniques that have been used for medical image analysis are explained for a better understanding of the differences between the discussed approaches. The description is divided into supervised—convolutional neural network (CNN) and recurrent neural network (RNN)—and unsupervised—auto-encoder (AE), stacked auto-encoder (SAE), variational auto-encoder (VAE), restricted Boltzmann machine (RBM), deep belief network (DBN), and generative adversarial network (GAN)—learning models. The training dataset of supervised models includes input data and the respective response values. From it, the supervised learning algorithm seeks to create a model that can make predictions of the response values for unseen data. Unlike supervised models, the training data of unsupervised models comprises examples without any corresponding target values (Hastie et al. 2001). First of all, as most deep architectures are based on ANNs, a brief explanation is given.

### 2.1 Artificial neural networks

A generic ANN is composed of one or more layers of neurons through which a signal is propagated. Feature vectors are fed to the first layer. Each feature is assigned a weight by each neuron of this layer, which outputs a value based on the response of an activation function to the weighted sum of all features (plus an optional bias). The last layer is the output layer. The training process involves adjusting the parameters: the weights and bias of the neurons are iteratively updated with the purpose of minimising a error between the obtained and expected output.

The simplest ANN is a perceptron, consisting of only one neuron layer which is a linear classifier, i.e., an algorithm that classifies input by separating two categories with a straight line. However, not all the classification tasks are linearly separable. By adding neuron layers, the model is referred to as a multi-layered perceptron (MLP) that is usually applied in scenarios where data is not linearly separable. The intermediate layers are known as hidden layers and can have an arbitrary number. The MLP is typically trained with a backpropagation algorithm.

When the network has many layers it is often called a deep neural network (DNN).

### 2.2 Convolutional neural network

Among the various DL models, CNNs have become a promising type of models for medical image analysis. Without the need to manually customise hand-crafted features, they can retrieve hierarchical features by elaborating high-level features from low-level ones. Although this technique was first reported in 1980 (Fukushima and Miyake 1982), it only gained relevance when Krizhevsky et al. (2012) won the ImageNet challenge in 2012, with the AlexNet architecture.

There are some widely used models, such as LeNet, AlexNet, VGG19, GoogLeNet and ResNet, just to name a few (Litjens et al. 2017). Fully convolutional networks (FCNs) by Long et al. (2015) become attractive as they allowed segmentation delineations to be created from any image size and solved the problem of low resolution outputs by the degradation of input image quality caused by successive convolution operations and pooling layers (Litjens et al. 2017). A fully connected layer is not always necessary and can be replaced with a convolution layer (Gu et al. 2018). In FCN, the fully connected layers are transformed as convolutions and the up-sampling is made using deconvolutional layers. Inspired by FCN, Ronneberger et al. (2015) proposed the widely used U-Net architecture that was later extended to a three-dimensional (3D) version, 3D U-Net (Çiçek et al. 2016). A full description of a usual CNN architecture can be found in Gu et al. (2018).

### 2.3 Recurrent neural network

RNNs use time-series information and have been applied in scenarios with sequential input and delayed decision, managing the data in a different manner than CNN. It processes and learns features of sequential input and outputs another sequence (Feng et al. 2017). As another family of ANNs, it is composed of input nodes, output nodes and hidden nodes, where connections between units form a cycle with a one way direction creating a dynamic temporal behaviour. In contrast to CNNs, which have a feed-forward network structure, RNNs have internal memory (Lee et al. 2017).

So far, supervised models have been discussed. The following described deep models are used for unsupervised feature representation learning.

### 2.4 Auto-encoder and stacked auto-encoder

AE are networks with the purpose of projecting the input into a lower dimensional subspace representing its dominant latent or compressed structure by the minimisation of the reconstruction error between the input and output values of the network. SAEs are used to implement the idea of *greedy layer-wise unsupervised learning* with the aim of improving the representational power by using the activation values of the hidden units of one AE as input to the next AE (Erhan et al. 2010). The different layers of the network represent different levels of information (Shen et al. 2017a). To restrain the AE from learning a superficial solution, the model can be trained to recreate the input from a degraded version with the addition of noise. This configuration is called the denoising AE (Litjens et al. 2017).

### 2.5 Variational autoencoder

VAE is a generative model that estimates the Probability Density Function (PDF) of the training data (Kingma and Welling 2014). VAEs are called “auto-encoders” only because the final training objective that derives from their setup does have an encoder and a decoder, and resembles a traditional autoencoder (Doersch 2016). Unlike sparse autoencoders, there are generally no tuning parameters analogous to the sparsity penalties. And unlike sparse and denoising autoencoders, we can sample directly from the distribution  $P(X)$  [without performing Markov Chain Monte Carlo (MC)].

The encoder is an ANN whose input is a datapoint  $x$ , and the output is a hidden representation  $z$ . The decoder is another ANN whose input is the representation  $z$ , and outputs the

parameters of the PDF. The loss function of the VAEs is the negative log-likelihood with a regularizer. One of the reasons that make VAEs is that they can be trained with stochastic gradient descent.

## 2.6 Restricted Boltzmann machine and deep belief network

As described by Fischer and Igel (2014), RBMs are a sub-type of Boltzmann machines. The latter is used to learn and adjust its parameters to fit the probability distribution to the training data. Thus it can be considered as a probabilistic graphical model, also known as Markov random fields (MRFs). The purpose is, based on samples, to learn significant aspects of an unknown target distribution providing the distribution hidden in the training data. After successful learning, an RBM can be used as a generative model that allows sampling from the learned distribution. RBMs are composed of an input layer with visible units that correspond to the components of an observation and a hidden layer that possesses the latent input representation with the dependencies between the components of observations. Each neuron is connected to all the neurons in the other layer but there are no connections between neurons in the same layer. A DBN is a multi-layer architecture, constructed by stacking multiple building blocks composed of RBMs, in a hierarchical manner and fine-tuning the resulting deep network to learn high-level representations (Shen et al. 2017a).

## 2.7 Generative adversarial networks

In short, GANs are composed of two parts, a Generator ( $G$ ) and a Discriminator ( $D$ ), competing among them to generate realistically looking images ( $G$ ) and distinguish between real and synthetically generated images ( $D$ ). This framework corresponds to a minimax two-player game (Goodfellow et al. 2014).  $G$  is set against an adversary  $D$ , which learns to determine whether a sample is from the model distribution or the data distribution. As an illustration, in Goodfellow et al. (2014) it is suggested to think of  $G$  as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while  $D$  is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles.

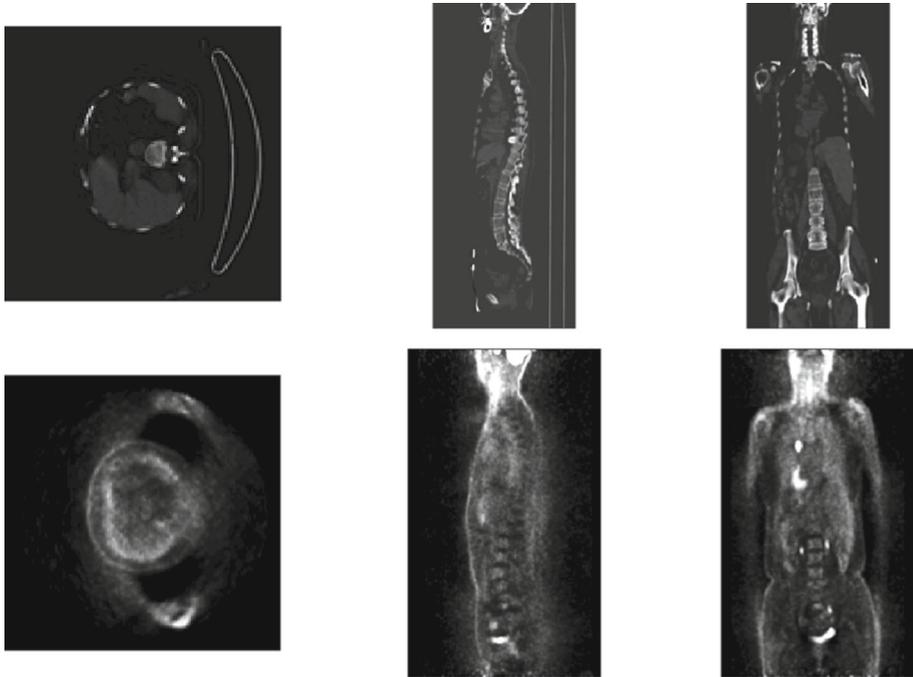
A variation to GANs, are the conditional GAN (cGAN). cGANs learn a mapping from observed image  $x$  and random noise vector  $z$ , to  $y$ . In this way, these networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping (Isola et al. 2017).

## 2.8 Medical imaging modalities

The imaging modalities considered in this review are following described. An illustration is given in Fig. 1.

### 2.8.1 Computed tomography

CT is a structural imaging procedure that integrates multiple X-ray projections taken from different angles by rotating the X-ray tube around the body to produce cross-sectional images of areas inside the body. Simultaneously, a detector array, at an opposite location of the X-ray tube, collects the transmission projection data, which is based on the principle that the

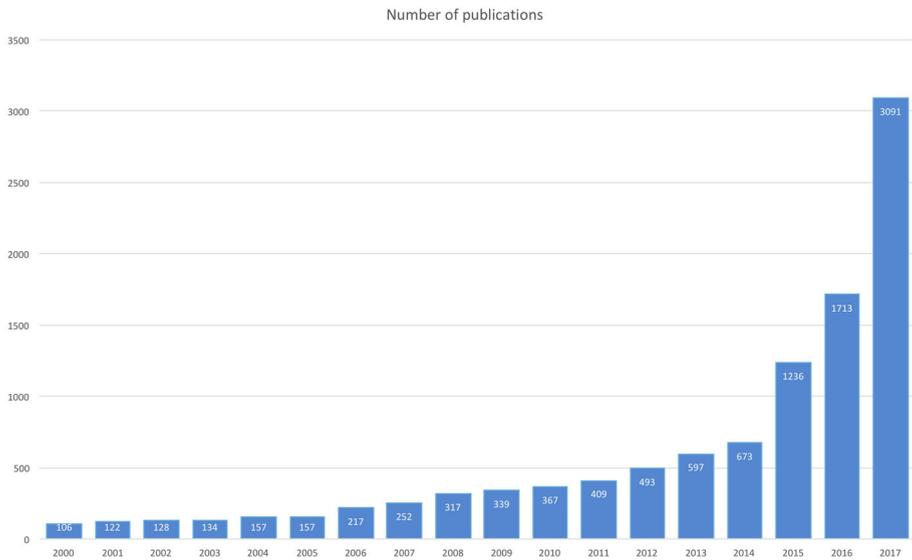


**Fig. 1** Example of a CT (top) and a PET (bottom) scans. Volumes are represented by an axis (left), a sagittal (middle) and a coronal (right) views

density of the tissue passed through by the X-ray beam can be measured by calculation of the attenuation coefficient (Bushberg and Boone 2011). In a CT slice, the grey levels correspond to X-ray attenuation that reflects the proportion of X-rays scattered or absorbed as passing through each voxel. X-ray attenuation is primarily a function of X-ray energy and the composition of the material being imaged (McCullough et al. 2015). There are limitations inherent to radiologic imaging studies that limit their diagnostic accuracy (ACC) such as the lack of functional information. Also, radiation exposure is a risk during CT examinations as it involves the attenuation of X-rays that interact by a combination of absorption events and scattering events and lead to the deposition of radiation dose in the patient (Brenner and Hricak 2010). However, the clinician chooses to perform CT exams when the benefits of the examinations far outweigh the radiation doses received (Lin 2010).

### 2.8.2 Positron emission tomography

PET is a nuclear imaging technique that enables the observation of metabolic processes via the injection of radiolabeled tracers, detection of the emitted radiation, and reconstruction of the radiolabeled tracer distribution. This technique is based on the detection of annihilation photons produced after positron emission from a radiolabeled tracer, which follows a specific molecular biological pathway (Prassopoulos and Efthymiadou 2016). The signalers are positron emitting nuclides and the overall tracer is administered as an intravenous injection, usually labelled with oxygen-15, fluorine-18, carbon-11, nitrogen-13 or gallium-68 (Wadsak and Mitterhauser 2010). The detached positron proceeds along the surrounding

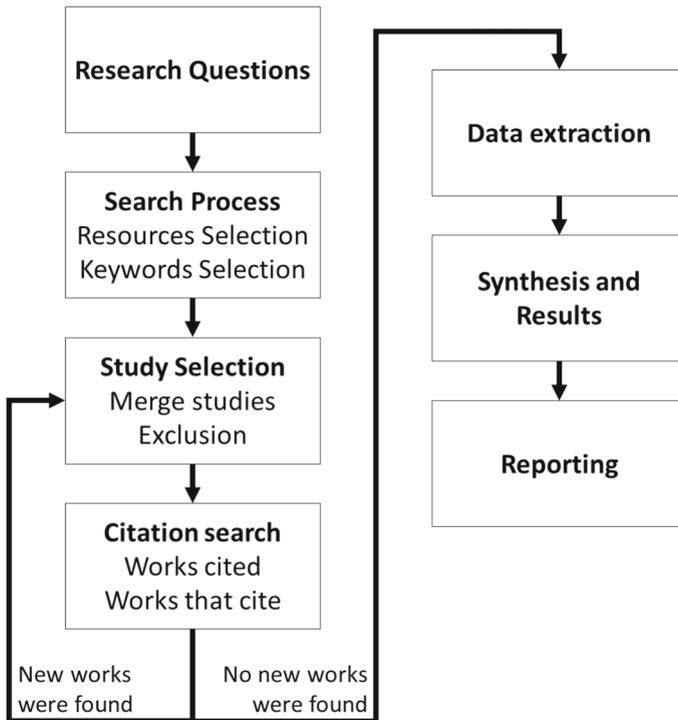


**Fig. 2** Distribution per year of papers related to works with DL over the period 2000–2017

material until it loses nearly all its energy. At this moment, the positron interacts with an electron and they annihilate each other by producing two gamma-ray photons emitted in opposite directions. Ideally, only the coincidence photons are registered (if the annihilation photons are detected within a limited time frame or coincidence detection) while random or scattered events are rejected (Townsend 2004). When photons reach the detectors, their energy is converted into light which is converted to electric signal. In the end, tomographic images are reconstructed through the application of algorithms and the biodistribution of the tracer is visualised. 18F-fluorodeoxyglucose (FDG) is the most common PET tracer due to its remarkable characteristics for physiological applications (Vaquero and Kinahan 2015). Its distribution reflects the distribution of glucose inside the body and thus has high sensitivity in detecting malignancies, and is routinely applied clinically for staging and monitoring treatment (Lin et al. 2015). Positron emission tomography (PET) imaging modality is characterised by the absence of identifiable anatomical structures caused by low contrast and resolution for precise anatomical localisation (Moses 2011). The spatial resolution is generally inadequate for anatomic localisation of pathology (Townsend 2008).

### 3 Overview

By performing a search on Web of Science Thomson Reuters website using “Deep Learning” as the query and selecting “article” as the type of document, it is readily seen that researchers have taken notice of DL approaches since the number of papers has been substantially increasing since 2000 (Fig. 2), reaching more than 3000 publications in 2017. This growth is mostly explained by the increasing number of application domains in which DL techniques have been playing an important role.



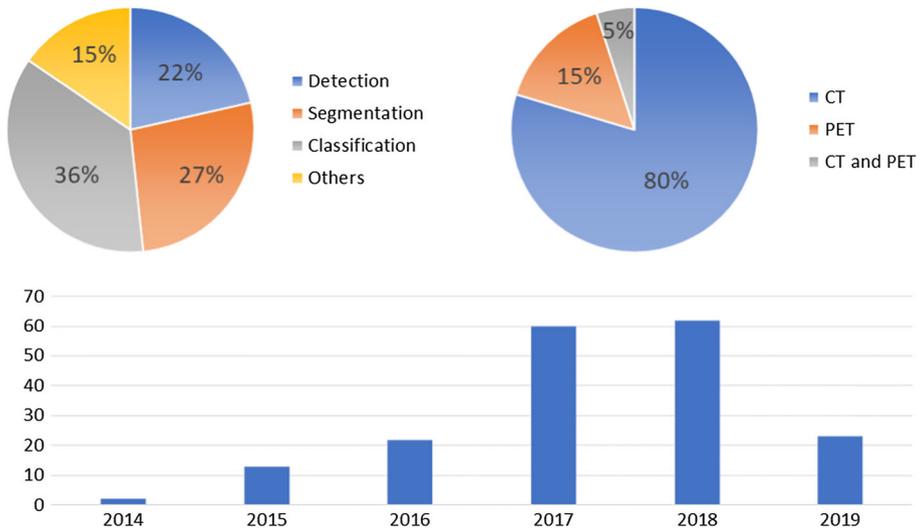
**Fig. 3** Literature review protocol

Figure 3 shows a flowchart of the protocol used in the Literature Review (Nazir et al. 2019; Asadi et al. 2019; Domingues et al. 2019). First, four main research questions were identified. These include:

- How is DL being used to analyse CT scans?
- How is DL being used to analyse PET scans?
- What are the benefits and key challenges of DL in medical imaging?
- How can DL contribute to aid specialists in their daily clinical practice?

Next, the sources to be consulted were selected. We have used four well-known databases, Web of Science, Scopus, PubMed, and Springer. The search strings included a combination of the keywords: DL, CNN, medicine, health, CT, PET and their derivations. The results were scanned by merging identical entries. They were then screened first with basis on the titles and abstracts. The selection criterion was to retrieve works that used DL techniques to analyse CT and/or PET scans. Moreover, only works written in English were considered. When pertinent, citations within the articles were added to the pool of papers. Works that cite the ones already in the pool were also verified. Having selected the papers of interest, they were carefully read in order to extract the main goal of the work, the anatomic interest, the methods used, the evaluation metrics and the characteristics of the dataset used in the experiments, the main results, and any other relevant information. This information was organised and the present document was written.

Some of the characteristics of the 182 reviewed papers can be seen in Fig. 4. It can be observed that most papers deal with a classification problem, followed by segmentation and



**Fig. 4** Statistics on the reviewed papers. Distribution per goal of the paper (top-left), distribution per type of exam (top-right), distribution per year (bottom)

then detection. Less attention has been paid to other types of problems. Concerning the data used, there is a very high prevalence of CT, followed by PET. To combine information from both types of images is still under-studied. The evolution per year of the selected papers mimics the one already illustrated in Fig. 2. The apparent decay in 2019 is only related to the fact that the search for the papers was made in the first trimester of 2019.

## 4 Literature review

This section describes the contributions of DL techniques to various generic tasks in medical image analysis: detection (Sect. 4.1), segmentation (Sect. 4.2), classification (Sect. 4.3), and others (Sect. 4.4), focusing on two different types of image exams: CT and PET.

The first three sections summarise works on each part of a typical computer-aided detection (CAD) pipeline. Detection works include the task of localisation of region of interests (ROIs) or volume of interests (VOIs). Segmentation refers to the extraction of a contour from a ROI or VOI. Classification is the categorisation into classes. The last section summarises other works, namely in the improvement of image quality (Sect. 4.4.1), reconstruction of dynamic PET (Sect. 4.4.2), extraction of time of flight (TOF) information from the detector signals (Sect. 4.4.3), image generation (Sect. 4.4.4), and the study of techniques that allow the application of a smaller radiation dose to the patient (Sect. 4.4.5).

### 4.1 Detection

Detection of objects, such as organs, landmarks or lesions in medical images, is an important pre-processing step in classification, for retrieval of medical images from databases, and may benefit radiological workflow (Domingues and Cardoso 2013). Some works with DL techniques applied with the purpose of detection of landmarks are revised next with CT,

PET and applications that use both CT and PET information. The main characteristics of the analysed studies are summarised in Tables 1, 2, 3 and 4.

#### 4.1.1 CT applications

A big plethora of works has already been published on detection using CT data. To keep this section from being too long, conference works are only summarised in Table 2. Other works and their summary is given in Table 1.

Liu et al. (2015) detect exophytic lesions, a subset of renal lesions, by (1) exploiting efficient belief propagation to segment kidneys, (2) establishing an intrinsic manifold diffusion on kidney surface, (3) searching for potential lesion-caused protrusions with local maximum diffusion response, and (4) exploring novel shape descriptors, including multi-scale diffusion response, with machine learning to classify exophytic renal lesions. Experimental results on the validation dataset with computed tomographic colonography (CTC) of 167 patients revealed that manifold diffusion significantly outperformed conventional shape features and resulted in 95% sensitivity with 15 false-positives (FPs) per patient for detecting.

Using a combination of full training and transfer learning, Shin et al. (2016) investigated the influence of different factors in deep CNNs on two specific CAD tasks, thoraco-abdominal lymph node (LN) detection and interstitial lung disease (ILD) classification. The thoraco-abdominal LN datasets were composed of 388 mediastinal LNs in 90 patient, and 595 abdominal LNs. The ILD dataset was composed of 905 image slices from 120 patients with six lung tissue types. They exploited different CNN architectures—CifarNet, AlexNet, OverFeat, VGG16 and GoogLeNet—with a variation of parameters and number of layers. Also, they explored the impact of dataset scales on performance. Finally, the benefits of transfer learning were investigated from non-medical to medical image domains. In their many experiments, 0.85 area under the curve (AUC) on the mediastinal LN detection and 0.92 on the abdomen LN detection was achieved. For the ILD, 0.53 and 0.79 of overall classification ACC were obtained with slice-level and patch-based classification, respectively. Their main conclusions were that CNN architectures with 8 to 22 layers were useful for CAD problems. Transfer learning from a large-scale annotated database to CAD problems proven to be worthwhile. The inferences made could be useful recommendations for the design of new CAD systems.

Wolterink et al. (2016) presented a method for automatic coronary artery calcification quantification in cardiac CT angiography. Their method made use of pairs of CNN to extract texture features from triplanar two-and-a-half-dimensional (2.5D) and volumetric 3D input patches. The dataset consisted of 250 cardiac CT exams divided into two sets. The first 50 exams were used to train the first part of the algorithm and the remaining 200 exams were used to train and evaluate the second part. The first CNN selected voxels likely to be coronary artery calcification that were further classified by the second CNN to distinguish between coronary artery calcification and negatives voxels. One difference was the purely convolutional nature with no pooling layers and fully connected layers were implemented as convolutions, allowing for fast voxel classification. In all the experiments, the fusion of 2.5D and 3D information outperformed their individual performance, with the best result of 71% lesion sensitivity and 0.48 FP error per scan.

For anatomical body parts identification on transversal slices, Yan et al. (2016) designed a two stage DL method. In the pre-train stage, a CNN was trained to extract discriminative patches from the training slices in a multi-instance way to exploit the local information, instead of using the entire slice. The second step consisted of a boosting stage using selected local patches. They performed a comparison study with both traditional classifiers—logistic

regression (LR) and support vector machine (SVM)—and other DL based approaches—CNN trained on the whole image, local patch-based CNN without boost, and local patch-based CNN boosted without additional non-informative class. The method was applied to 7486 transversal slices of CT scans divided into 12 anatomical classes and the proposed method achieved the best results with 92.23% of classification ACC. To investigate the capability of CNN to discover discriminative features, the learned hidden activation on each intermediate layer was extracted and further used to train a SVM classifier. The best F1-score on testing set was 0.88, concluding that features of higher layers were better for the classification task.

Belharbi et al. (2017) aimed at spotting the third lumbar vertebra (L3) slice within a whole body CT scan. They compared the performance of a fully trained CNN composed of four convolutional layers followed by a fully connected output layer with pre-trained CNNs over ImageNet dataset. The dimension of the input was converted from 3D to two-dimensional (2D) using a frontal Maximum Intensity Projection. All the models were tested on 642 CT exams and an extra set of 43 CT scans, annotated by the same radiologist and 3 other experts. Pre-trained models achieved better results (AlexNet =  $2.45 \pm 2.42$ , VGG16 =  $1.82 \pm 2.32$ , VGG19 =  $1.83 \pm 1.83$ , and GoogLeNet =  $2.54 \pm 4.22$ ) concerning the average error measured in slices when compared with the homemade CNN approach ( $2.78 \pm 2.48$ ). On the other hand, the homemade CNN required only 4.46 s per CT scan to find the L3 slice what could allow the integration of the system into daily clinical routines while VGG16 required an average of 13.28 s per CT scan.

Similarly, de Vos et al. (2017) developed an automatic system for localisation of anatomical structures in 3D images as a 2D detection task, which means that the structures were first detected on 2D images slices extracted from orthogonal planes. With the use of an independent CNNs, the anatomical targets were detected on each axial, coronal and sagittal slices. Next, 3D bounding boxes were created by integrating the result of the CNN in all slices, summing their outputs. A bounding box network (BoBNet) architecture was used consisting of eight convolutional layers. Three series of CT images were used—200 chest low-dose CT (LCT), 100 cardiac CT angiography, and 100 abdomen CT scans. The authors claimed to have obtained an accurate method to localise multiple structures, particularly the left ventricle with 4.5 mm for centroid distance (CD) between automatic and reference bounding box, 3 mm for wall distance between automatic and reference bounding box walls, and 2.5 s of processing time.

Vivanti et al. (2017) describe a method for automatic detection and segmentation of tumours in longitudinal liver CT studies and for liver tumours burden quantification. The method integrates information from the scans, the baseline known tumours delineations, and a tumour appearance prior model in the form of a global CNN. Its outputs are the new tumours segmentation in the follow-up scan, the tumour burden quantification in both scans, and the tumour burden change. Experimental results on 246 tumours, 97 of which were new tumours, from 37 longitudinal liver CT studies with radiologist approved ground-truth segmentation, yields a true-positive (TP) new tumours detection rate of 86 versus 72% with stand-alone detection, and a tumour burden volumetric overlap error (VOE) of 16%. The authors claim that, unlike other deep learning-based methods, this method does not require large tagged training sets.

Wang et al. (2017b) aims at detecting subcutaneous fat areas (SFA) and visceral fat areas (VFA). The proposed CAD framework consists of two steps with two CNNs namely, Selection-CNN and Segmentation-CNN. The first CNN was trained using 2240 CT slices to select abdominal CT slices depicting SFA and VFA. The second CNN was trained with 84,000 pixel patches and applied to the selected CT slices to identify fat-related pixels and assign them into SFA and VFA classes. Comparing to the manual CT slice selection and fat

pixel segmentation results, the ACC of CT slice selection using the Selection-CNN yielded 95.8%, while the ACC of fat pixel segmentation using the Segmentation-CNN was 96.8%.

Different loss functions can be applied efficiently in different applications and it is therefore important to study their influence on the results, namely in the way of penalising straightforward examples and the robustness to labelling noise. Barbu et al. (2018) compared the effect of four loss functions for training deep CNNs in LN detection. The dataset was formed by 176 CT volumes containing 388 mediastinal LNs and 595 abdominal LNs. The four loss functions considered were the logistic loss, the differentiable approximation of the SVM loss, the Lorenz loss and a modified version of the tangent loss. LN candidates were obtained using a random forest (RF) approach and a multi-label fusion from datasets containing mediastinal and abdominal LNs. As input to the CNN, 36 triplets of orthogonal cross sections were used, extracted at random directions and scales. The authors concluded that the commonly used logistic loss based on softmax function may not always be the most appropriate choice. The Lorenz loss obtained the best performance with an improvement of 3% in detection rate at 3 FP per scan compared to the logistic loss function.

Blau et al. (2018) address the detection of simple renal cysts as an incidental finding in a real clinical setting. A combined 3D distance map of the kidneys and surrounding fluids provides initial candidates for cysts. A CNN then classifies the candidates as cysts or non-cyst objects. Performance was evaluated on 52 abdominal CT containing over 70 cysts. By setting the minimal cyst diameter to 10 mm, the algorithm detected 59/70 cysts (TP rate = 84.3%) while producing an average of 1.6 FP per case.

Tomita et al. (2018) present and evaluate a system that can detect incidental osteoporotic vertebral fractures (OVFs) in chest, abdomen, and pelvis CT examinations. The system leverages a CNN to extract radiological features from each slice in a CT scan. These extracted features are processed through a feature aggregation module to make the final diagnosis for the full CT scan. Several techniques are explored for feature aggregation, including the use of a long short-term memory (LSTMs) network. The system was trained and evaluated on 1432 CT scans, comprised of 10,546 2Ds images in sagittal view. It achieved an ACCs of 89.2% and an F1-score of 90.8%. The authors claim that the results match the performance of practising radiologists.

Diniz et al. (2019) propose a computational methodology capable of detecting spinal cord in planning CT images. The pipeline consists of the following four steps: (1) images acquisition, (2) initial segmentation, (3) candidates segmentation and (4) candidates classification. Initial segmentation is made with an adaptive template matching; candidate segmentation with an intrinsic manifold simple linear iterative clustering (IMSLIC); and candidate classification with a CNN. The methodology was applied on 36 planning CT images provided by The Cancer Imaging Archive (TCIA), and achieved an ACC of 92.55%, specificity of 92.87% and sensitivity of 89.23% with 0.065 of FPs per image.

Gerard et al. (2019) propose a supervised discriminative learning framework for simultaneous feature extraction and classification. The proposed framework, called FissureNet, is a coarse-to-fine cascade of two CNNs. The coarse-to-fine strategy alleviates the challenges associated with training a network to segment a thin structure that represents a small fraction of the image voxels. FissureNet was evaluated on a cohort of 3706 subjects with inspiration and expiration CT scans from the COPDGene clinical trial and a cohort of 20 subjects with 4-dimensional computed tomography (4DCT) scans from a lung cancer clinical trial. On both datasets, FissureNet showed superior performance compared to a deep learning approach using the U-Net architecture and a Hessian-based fissure detection method in terms Precision–Recall AUC. The overall Precision–Recall AUC for FissureNet on the COPDGene dataset was 0.980 and on the lung cancer dataset was 0.966.

Ghesu et al. (2019) reformulate the detection problem as a behaviour learning task for an artificial agent. The modeling of the anatomy appearance is coupled with the object search in a unified behavioural framework, using the capabilities of deep reinforcement learning and multi-scale image analysis. In other words, an artificial agent is trained not only to distinguish the target anatomical object from the rest of the body but also how to find the object by learning and following an optimal navigation path to the target object in the imaged volumetric space. The approach is evaluated on 1487 CT volumes from 532 patients. The authors claim to successfully detect several anatomical structures with no failed cases, while also improving the detection-speed, achieving unmatched real-time performance.

Xie et al. (2019a) develop a pulmonary nodule detection framework with 2D CNN. Firstly, the structure of Faster R-CNN is adjusted with two region proposal networks and a deconvolutional layer to detect nodule candidates, and then three models are trained for three kinds of slices for later result fusion. Secondly, a boosting architecture based on 2D CNN is designed for FP reduction. The misclassified samples are kept for retraining a model which boosts the sensitivity. Finally, the results of these networks are fused to vote out the final result. Experiments are conducted on LUNA16 dataset, and the sensitivity of nodule candidate detection achieves 86.42%. For the FP reduction, the sensitivity reaches 73.4% and 74.4% at 1/8 and 1/4 FPs/scan, respectively.

Table 1 resumes the characteristics of the previously discussed works and Table 2 resumes the characteristics of conference works with the aim of detecting landmarks on CT applications. The dataset is indicated with the sub-type of CT acquisition as well as the number of instances.

**Conclusions** Under the application of DL techniques concerning CT modality, the detection task has been mostly approached as a full training problem using CNNs with 1D, 2D or 3D convolutions. Fewer works have integrated these techniques as transfer learning and deep feature extractors. The preponderance of the use of CNN can be explained by the fact that this architecture is known to capture relationships between adjacent pixels, which is fundamental in detecting landmarks. In this sense, other techniques should be explored. For instance, a AE or a stacked denoising auto-encoder (SDAE) which could be useful to extract latent-space representation encoded into the ROI of detection.

From the Table 1, it is observed that DL techniques have been applied in a wide range of anatomical applications. Not only with the aim of detecting one particular structure, but also to detect several structures in one system which significantly increases the complexity of the proposed problem. However, the detection task has been simplified in many works to a 2D approach with the CT already restricted to the zone of interest, rather than a whole-body 3D approach. A variety of evaluation metrics for landmarks detection have been used. With high FP rates, we can verify that automatic detection methods have not evidenced sufficiently robust results for clinical use.

Detection is highly susceptible to variation in medical images caused by differences in image acquisition, anatomy, and pathology among subjects. Real applications require robust detection systems as missing a lesion or landmarks on a scan can have serious consequences for both the patient and the clinician. One way to achieve that could be to increase the amount of training data as well as the inclusion of heterogeneity on the training data with images obtained in different acquisition and patient conditions.

**Table 1** Main characteristics of the studies reviewed in Sect. 4.1.1: CT detection

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Liu et al. (2015)	Kidney	167 CTc	Sensitivity	95% at 1.5 FPs/patient
Shin et al. (2016)	Thoraco-abdominal LN and ILD	90 + 86 CT volumes, and 905 CT slices	AUC	0.85 for LN detection, 0.95 for ILD classification
Wolterink et al. (2016)	Coronary artery calcification	250 cardiac scans	Sensitivity, FPs per scan	71%, 0.48
Yan et al. (2016)	Several body-parts	7489 transversal slices	ACC, F1	92.23%, 0.88
Belharbi et al. (2017)	L3	642 + extra 43 whole body CT volumes	Localisation error of slices	1.91 ± 2.69 mm
de Vos et al. (2017)	Several anatomical structures	200 chest LCT, 100 cardiac CT angiography, and 100 abdomen CTs	Wall distance, CD, Processing time	3.0 ± 3.7 mm, 4.5 ± 3.4 mm, 2.5 s
Vivanti et al. (2017)	Liver	246 tumors from 37 CTs	AUC, precision, sensitivity	0.91, 0.87, 0.86
Wang et al. (2017b)	SFA and VFA	2240 CT slices	ACC	95.8%
Barbu et al. (2018)	LN	176 scans	Increase in detection rate, FPs per scan	3%, 3
Blau et al. (2018)	Kidneys	52 abdominal CT containing > 70 cysts	Sensitivity	84.3% at 1.6 FP/case
Tomita et al. (2018)	OVF	10,546 2Ds images in sagittal view of 1432 CTs	ACCs, F1	89.2%, 90.8%
Dimiz et al. (2019)	Spinal cord	36 images from TCIA	ACC, specificity, sensitivity	92.55%, 92.87%, 89.23% at 0.065 FPs/image
Gerard et al. (2019)	Lung	CTs of 3706 subjects and 4DCTs of 20 subjects	Precision–Recall AUC	0.980 for CTs and 0.966 for 4DCTs
Ghesu et al. (2019)	8 anatomical landmark points	1487 CTs from 532 patients	ACCs, speed	4.19, 0.061
Xie et al. (2019a)	Lung	LUNA16	Sensitivity at 0.125, 0.25, 0.5, 1, 2, 4, 8 FPs/patient, Competition Performance Metric (CPM)	0.734, 0.744, 0.763, 0.796, 0.824, 0.832, 0.834, 0.790

**Table 2** Main characteristics of the conference works on CT detection

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Roth et al. (2014)	LNs	90 CT volumes with 388 mediastinal LNs and 86 patients with 595 abdominal LNs	Sensitivity	70%/83% at 3 FPs/vol, 84%/90% at 6 FPs/vol
Shen et al. (2015a)	Vertebrae centres	10 chest CT images (SpineWeb dataset)	Mean pixel distance between the predicted centres to the ground truth centres	1.3987 mm
Zheng et al. (2015)	Carotid artery bifurcation	455 head-neck CT volumes	Mean error, Detection speed per volume	2.64 mm, 1 s
Golan et al. (2016)	Lung	1280 thoracic scans [Lung Image Database Consortium (LIDC)-Image Database Resource Initiative (IDRI)]	Sensitivity	78.9% at 20 FPs/scan, or 71.2% at 10 FPs/scan
Liu et al. (2016)	Colitis	260 images 26 patients	Sensitivity	0.85 at 1 FP/image
Cha et al. (2017b)	bladder	57 patients with 71 biopsy-proven lesions	Sensitivity	84.2% at 3.7 FPs/case
Hamidian et al. (2017)	Pulmonary nodules	509 cases (LIDC-IDRI)	Sensitivity	80% at 22.4 FPs/case, and 95% at 563 FPs/case
Kalinovsky et al. (2017)	Lung	338 scans of tuberculosis patients	AUC	0.969
Keshavamurthy et al. (2017)	Brain	409 scans of patients with traumatic brain injury (TBI)	ACC, sensitivity, specificity	92.55%, 91.15%, 93.45%
Lisowska et al. (2017)	Brain	Scans for 122 patients with suspected acute ischaemic stroke	Receiver Operating Curve (ROC) AUC, Precision-Recall AUC	0.996, 0.563
Liu et al. (2017)	Pelvic artery calcification	500 images from 20 scans	Precision, sensitivity, FP	77.4%, 85%, 1
Näppi et al. (2017)	Serrated polyps	101 CTC	Sensitivity	93% at 0.8FPs/patient
Umehara et al. (2017)	Polyps	154 CTC	Sensitivity	93.5% at 3.9FPs/patient
Chen et al. (2018)	Polyps	144 CTC	Sensitivity	0.97 at 20.3FPs/scan
Dormer et al. (2018)	Heart	Chest CT for 12 patients	AUC, ACC	0.840, 78.9%
Grewal et al. (2018)	Brain	77 scans	ACC	81.82%
Hu et al. (2018)	Lung	LUNA16	Sensitivity	94.3%
Oda et al. (2018)	Mediastinal lymph node	45 contrast-enhanced chest CTs	Sensitivity	95.5% at 16.3FPs/scan

### 4.1.2 PET applications

All of the works found concerning detection on PET consider the identification of normal FDG uptake and excretion (sFEPUs) (Afshari et al. 2018; Deng et al. 2017; Bi et al. 2017b). Identification of normality is important since normal cases can be screened-out from the process, leaving the specialists more time to focus on more difficult cases (Bessa et al. 2014).

sFEPUs can be incorrectly identified as anomalies. Having this in mind, Bi et al. (2017b) develop a method for the detection and classification of these regions. A multi-scale superpixel-based encoding (MSE) is used to group sFEPUs into larger regions. Feature extraction is then made via domain transferred CNNs. These regions are then classified into one of the sFEPUs classes (brain, bladder, heart, left kidney, right kidney, and other hypermetabolic fragments) using class-driven feature selection and classification (CFSC). Experiments on 40 whole-body lymphoma PET/CT studies show that the method achieves an average F-score of 91.73%.

The work (Deng et al. 2017) has as goal the detection of lymphomas. The pipeline has two main blocks, one for the detection of lymphomas, another for the detection of sFEPUs. The sFEPUs acts as a FP step and thus its results are subtracted from the lymphoma detection results. Both detection blocks have the same architecture consisting in a set of five FCN-8 networks in parallel. Manually slices of the abdomen and chest of 30 PET images, 10 of each from lymphoma patients are used in the experiments. Identification performance is 97.10 for the right kidney; 96.68 for the left kidney; 94.80 for the bladder; 91.59 for the liver and 83.56 for other high uptake areas. Interestingly, no results are given for the detection of lymphomas.

Afshari et al. (2018) are also interested in localising normal active organs. The method adapts the deep network architecture of YOLO (Redmon et al. 2016) to detect multiple organs in 2D slices and aggregates the results to produce semantically labelled 3D bounding boxes. Evaluation is made on 479 FDG-PET scans of 156 patients achieving an average organ detection precision of 75–98%, recall of 94–100%, average bounding box CD of less than 14 mm, wall localisation error of less than 24 mm and a mean Intersection Over Union (IOU) of up to 72%.

Table 3 summarises the main characteristics (anatomic interest, dataset, and best performance results) of the analysed study.

**Conclusions** The importance of sFEPUs detection is clear since they may cause a high prevalence of FP detections. It would be interesting, however, to study the application of DL to the detection of lesions or other structures using PET only.

### 4.1.3 CT and PET applications

Two works were found that combine information from both CT and PET. Both works use CNNs, but while in Teramoto et al. (2016), it is used for feature extraction only, in Xu et al. (2018) the full pipeline is DL. The third work, Kumar et al. (2018) also uses a full DL for the detection of ROIs. More detail on each work follow next and are summarised in Table 4.

Teramoto et al. (2016) work aimed at developing an automated scheme for the detection of lung tumours integrating PET and CT data with a multi-stream CNN. In their first study, they achieved a high sensitivity accompanied by a high number of FPs, 17 per case. After applying an active contour model (ACM) they improved their system, achieving a sensitivity of 90.0% with 9.8 FPs/case. With this work, they wanted to further improve their technique by lowering the FPs rate using a CNN. The overall implementation scheme is described next. First, nodule

**Table 3** Main characteristics of the studies reviewed in Sect. 4.1.2: PET detection

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Bi et al. (2017b)	sFEPU	40 whole-body lymphoma PET-CT studies	F-score	91.73%
Deng et al. (2017)	Lymphomas	Slices from 30 PET images	Identification performance	97.10 for the right kidney; 96.68 for the left kidney; 94.80 for the bladder; 91.59 for the liver and 83.56 for other high uptake areas
Afshari et al. (2018)	sFEPU	479 FDG-PET scans of 156 patients	Precision, recall, bounding box CD, wall localisation error, mean IOU	75–98%, 94–100%, < 14 mm, < 24 mm, 72%

**Table 4** Main characteristics of the studies reviewed in Sect. 4.1.3: CT and PET detection

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Xu et al. (2018)	Multiple myeloma bone lesion	12 clinical 3D whole body PET/CT	Sensitivity, specificity, precision, dice	73.50%, 99.59%, 72.46%, 72.98%
Kumar et al. (2018)	Lungs, mediastinum, tumors, and background	50 FDG-PET/CT scans of patients with NSCLC	Precision, sensitivity, specificity, and ACC	78.16, 98.00, 99.26, 99.23

candidates were located using CT and PET images separately using different algorithms: for CT the region was enhanced using an ACM followed by segmentation through thresholding and labelling and PET images were binarized to detect regions of increased uptake. The two resulting images were then merged and combined using the logical OR function. In order to eliminate FPs in the initial candidates while maintaining the value of TPs, an ensemble method using a multi-step classifier consisting of two features extractors—shape from CT and metabolic feature from PET analysis and the other by a CNN followed by a two-step classifier rule and SVMs. The architecture of the CNN consisted of three convolution layers, three pooling layers, and two fully connected layers. The authors evaluated the detection performance using 104 PET/CT images obtaining a sensitivity of detection of 90.1%, with 4.9 FPs/case. The results showed that the CNN extracted features are valuable for FP reduction.

Xu et al. (2018) implemented two CNNs, V-Net and W-Net, for lesion prediction and segmentation with the intention of detecting a multiple myeloma bone lesion in 3D whole body data. The V-Net architecture aimed at performing volumetric image segmentation through 3D convolutions both on CT and PET images. To combine extracted features from both modalities, they cascaded two V-NETs to form a W-Net. The method was evaluated on 12 clinical PET/CT data. The results demonstrated that the W-Net architecture provided best results (sensitivity = 73.50%, specificity = 99.59%, precision = 72.46%, dice = 72.98%) for segmentation and lesion detection with multimodal information when compared with the V-Net architecture tested on each modality independently. Furthermore, they compared DL techniques with traditional machine learning methods such as RF, k-nearest neighbours (kNN), and SVM with the first approach outperforming the classical ones that confirmed the advantage of DL approaches in handling lesion heterogeneities (sensitivity = 89.71%, specificity = 99.68%, precision = 88.82%, dice = 89.26%). This study also validated the advantage of the combination of PET and CT in providing discriminative features for the characterisation of the disease without manually extracting handcrafted features.

Kumar et al. (2018) present a CNN that learns to fuse complementary anatomical and functional data from PET/CT in a spatially varying manner, for the detection of different ROIs. The novelty is the ability to produce a fusion map that explicitly quantifies the fusion weights for the features in each modality. To achieve this, the CNN comprises four main components: two encoders (one for each modality), one co-learning and fusion component, and a reconstruction component. The used dataset comprises 50 FDG-PET/CT scans of patients with pathologically proven nonsmall cell lung cancer (NSCLC). The findings from region detection (lungs, mediastinum, tumors, and background) experiments demonstrate that the approach achieves a significantly higher ACC ( $p < 0.05$ ) than several baseline CNN-based methods for multi-modality image analysis.

Table 4 summarises the main characteristics (anatomic interest, dataset, and best performance results) of the analysed studies.

**Conclusions** Two of the analysed studies compared traditional techniques with CNN-based methods, proving the added value of DL. The main characteristics can be consulted in Table 4. The application of DL techniques that profit from both PET and CT information is still an under-studied topic. There is a need to continue to explore this context with the development of more works.

## 4.2 Segmentation

Image segmentation consists of dividing the original image into regions. The pixels that belong to a certain region have similar attributes, that is, pixels with the same label own

similar features. The purpose of medical image segmentation is to obtain specific anatomical structures (Martins et al. 2007) and/or find ROIs/VOIs such as tumours or lesions by the identification of the contour or the interior region (Litjens et al. 2017; Vial et al. 2018). Data mining techniques have been integrated to optimise basic methods based on grey level and textural features (Sharma et al. 2010).

Recently, DL, and in particular CNN, has proven to be effective in segmentation tasks being the most common topic of papers in the field of DL applied to medical imaging (Litjens et al. 2017). Class probabilities are assigned for each region, generating label probability maps. In this context, the segmentation task is seen as the distinction of a group of pixels/voxels which lie on the contour or interior of the object(s) of interest.

Some works with DL techniques applied through the segmentation workflow are revised next, considering CT, PET and both CT and PET modalities. The main characteristics of the analysed studies are summarised in Tables 5, 6, 7 and 8.

#### 4.2.1 CT applications

The development of segmentation techniques for CT data is very widespread. For this reason, conference papers are only gathered on Table 6, while other works are further described next and summarised in Table 5. Furthermore, we start by presenting single class works, followed by multi-class classification techniques.

##### Single-class classification

Cha et al. (2016) explored the application of a deep network for bladder segmentation. Using ROIs, a CNN was trained to distinguish the inside and the outside of the bladder through the likelihood estimation of a region being located inside or outside the bladder. That resulted in a likelihood map to guide the level set segmentation. After the application of thresholding and hole-filling methods to generate the initial contour of the bladder, 3D and 2D level sets refined the initial segmentation propagating the surface using cascading level sets. Their method was evaluated using 173 CT urography cases hand-segmented reference contours. The deep method based likelihood achieved a better segmentation compared to traditional approaches based on Haar features.

Dou et al. (2017) present a 3D FCN equipped with a 3D deep supervision mechanism, called 3D DSN. 3D DSN is capable of conducting volume-to-volume learning and inference and the 3D deep supervision mechanism can cope with the optimisation problem of gradients vanishing or exploding when training a 3D deep model, accelerating the convergence speed and simultaneously improving the discrimination capability. Such a mechanism is developed by deriving an objective function that directly guides the training of both lower and upper layers in the network, so that the adverse effects of unstable gradient changes can be counteracted during the training procedure. A fully connected conditional random field (CRF) model is employed as a post-processing step to refine the segmentation results. Validation was made on two datasets: (i) liver segmentation from CT and (ii) whole heart and great vessels segmentation from MRI. The authors claim to have achieved competitive segmentation results with a much faster speed.

Farag et al. (2017) present a bottom-up approach for pancreas segmentation. The method generates a hierarchical cascade of information propagation by classifying image patches at different resolutions and cascading (segments) superpixels. The system contains four steps: (1) decomposition of slice images into a set of disjoint boundary-preserving superpixels; (2) computation of class probability maps via dense patch labelling; (3) superpixel classification by pooling both intensity and probability features to form empirical statistics in cascaded

**Table 5** Main characteristics of the studies reviewed in Sect. 4.2.1: CT segmentation

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Özsavaş et al. (2014)	Lungs, trachea/main bronchi, and spinal canal	10 patients with lung cancer and 10 thoracic CT scans from LIDC	Volume overlap ratio, ASD	94.30 ± 3.93%, 0.28 ± 0.21 mm (lung) and 99.11 ± 0.26%, 0.89 ± 0.32 mm (LIDC)
Cha et al. (2016)	Bladder	173 CT urography volumes	Volume intersection ratio, percent volume error, absolute volume error, minimum distance, IOU	81.9% ± 12.1%, 10.2% ± 16.2%, 14.0% ± 13.0%, 3.6 ± 2.0 mm, 76.2% ± 11.8%
Dou et al. (2017)	Liver	MICCAI-SLiver07	VOE, RVD, ASD, RMSD, time maximum symmetric surface distance (MSD)	5.42, 1.75, 0.79, 1.64, 33.55, 1.5 min
Farag et al. (2017)	Pancreas	80 CTs	Dice, IOU	70.7%, 57.9%
Fechter et al. (2017)	Oesophagus	50 CTs	Dice, ASD, Hausdorff	0.76 ± 0.11, 1.36 ± 0.90 mm, 11.68 ± 6.80
Huang et al. (2017)	Osteosarcoma	405 CTs	Dice, sensitivity, Hammoede, F1	87.80%, 86.88%, 19.81%, 0.908
Ibragimov et al. (2017)	PV	72 CTs	Dice, ASD	0.83, 1.08 mm
Ibragimov and Xing (2017)	Head and neck	50 CTs	Dice (%)	87.0 (Spinal cord), 89.5 (Mandible), 76.6 (Parotid left), 77.9 (Parotid right), 69.7 (Submandibular left), 73.0 (Submandibular right), 85.6 (Larynx), 69.3 (Pharynx), 88.4 (Eye globe left), 87.7 (Eye globe right), 63.9 (Optic nerve left), 64.5 (Optic nerve right), 37.4 (Chiasm)
Men et al. (2017)	Rectal cancer	278 patients	Dice	87.7% (CTV), 93.4% (Bladder), 92.1% (Left femoral head), 92.3% (Right femoral head), 65.3% (Intestine), 61.8% (Colon)

Table 5 continued

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Sun et al. (2017a)	Liver	3Dircadb and JDRD	VOE, RVD, ASD, RMSD, MSSD	$15.6 \pm 4.3\%$ , $5.8 \pm 3.5\%$ , $2.0 \pm 0.9\%$ , $2.9 \pm 1.5$ mm, $7.1 \pm 6.2$ mm (3Dircadb) and $8.1 \pm 4.5\%$ , $1.7 \pm 1.0\%$ , $1.5 \pm 0.7\%$ , $2.0 \pm 1.2$ mm, $5.2 \pm 6.4$ mm (JDRD)
Yuan (2017)	Liver	LITS	Dice	0.963 (liver) and 0.657 (tumour)
Zhou et al. (2017a)	19 structures in the human torso	240 CT volumes	ACC	87.9%
Ait Skourt et al. (2018)	Lung	LIDC-IDRI	Dice	0.9502
Chmelik et al. (2018)	Spine	31 CTs involving 626 vertebrae	Sensitivity, specificity, ACC, AUC, Youden's index	0.71, 0.88, 0.80, 0.59 (Lytic) and 0.74, 0.82, 0.78, 0.56 (Sclerotic)
Gibson et al. (2018)	Abdomen	90 subjects	Dice, boundary distance, Hausdorff	0.63–0.96, 0.8–2.5, 2.4–15.0
Jackson et al. (2018)	Kidney	89 patients	Dice	0.91 (right kidney), 0.86 (left kidney)
Lustberg et al. (2018)	Lung	20 stage I–III NSCLC patients	Dice, Hausdorff	90%, 1.5 cm
Öman et al. (2019)	Brain	60 patients	Sensitivity, specificity, AUC, dice	0.93, 0.82, 0.93, 0.61
Trullo et al. (2019)	OARs	11,084 slices from 60 CTs	Dice, MSE	0.55, 3.06 (esophagus), 0.66, 32.97 (heart), 0.63, 6.93 (trachea), 0.76, 4.29 (aorta)
Xu et al. (2019)	Lung	201 subjects	Dice, Self-adaptability	96.71, 100
Zhu et al. (2019)	Lung	66 chest and abdominal CTs	Dice, ASD, Hausdorff	0.87–0.95, 1.8–3.8 mm, 7.9–11 mm (heart, lungs and liver), and 0.71–0.79, 1.2–2.2 mm, 4.0–7.9 mm (spinal cord and oesophagus)

**Table 6** Main characteristics of the conference works on CT segmentation

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Roth et al. (2015a)	Pancreas	82 abdominal CTs	Dice	68% ± 10%
Roth et al. (2015c)	Pancreas	82 abdominal CTs	Dice	71.8% ± 10.7%
Christ et al. (2016)	Liver and lesions	15 CTs with hepatic tumours (3Dircaadb dataset)	Dice	94.3%
Dou et al. (2016)	Liver	30 contrast-enhanced CTs (MICCAI-SLiver07)	VOE, RVD, ASD, RMSD, MSD	5.42%, 1.75%, 0.79 mm, 1.64 mm, 33.55 mm
Gao et al. (2016)	5 classes of lung tissue	128 CTs (ILD dataset)	ACC	92.8%
Zhou et al. (2016)	Anatomical structures	240 CTs	ACC	89 %
Gordon et al. (2017)	Bladder	94 CT urography	Volume intersection	87.6 ± 7.6% (inner wall) and 87.2 ± 9.3% (outer wall)
Humpire Mamani et al. (2017)	Six organs	553 thorax abdomen CTs	Dice	0.95 and 0.95 (lungs), 0.59 and 0.58 (kidneys), 0.83 (liver) and 0.63 (spleen)
Jin et al. (2017)	Liver	20 CTs (3Dircaadb dataset)	Dice, VOE, RVD, ASD, MSD	95.7, 6.5, 1.0, 1.2, 18.3
Ma et al. (2017)	Prostate	92 CTs	Dice	86.80%
Patel et al. (2017)	Brain	Non-contrast CT of eight adult patients	Dice, mean absolute volume difference, prediction time	0.87 ± 0.01, 4.77 ± 2.70%, 68 s
Shi et al. (2017)	Prostate	CT for 22 different patients	Dice, ASD, CD-x, CD-y, CD-z	0.888, 1.64, 1.03, 0.97, 0.65
Sun et al. (2017b)	Liver	3Dircaadb and JDRD databases	VOE, RVD, ASD, RMSD, MSD	14.0%, 5.8%, 2.0%, 2.9mm, and 7.1mm (3Dircaadb); 10.1%, 5.2%, 1.7%, 2.1mm and 6.1mm (JDRD)
Trullo et al. (2017a)	Oesophagus	30 thoracic CTs	Dice	0.72 ± 0.07

Table 6 continued

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Trullo et al. (2017b)	Multiple Thoracic Organs	30 CTs	Dice	0.69 (esophagus), 0.90 (heart), 0.87 (trachea), 0.98 (aorta)
Xu and Liu (2017)	Lung	20 VOEs (LIDC-IDRI)	Dice	0.754
Zhang et al. (2017)	Liver	3Dircadb	VOE, RVD, ASD, MSD, dice	6.97, 0.73, 0.84, 6.32, 95.27
Zhou et al. (2017b)	Anatomical structures	240 CTs	IOU, Sensitivity	99%, 88% (left kidney), 84%, 65% (pancreas)
Işgum et al. (2018)	Vertebrae	15 spine CTs (MICCAI CSI 2014 workshop challenge)	Dice, mean surface distance	0.948, 0.29 mm

RF frameworks; and (4) connectivity based post-processing. Dense image patch labelling is conducted using two methods: RF classification on image histogram, location and texture features; and CNN classification, on larger image windows. The method is evaluated on a data set of 80 volumes, achieving a Dice coefficient of 70.7% and IOU of 57.9%.

Fechter et al. (2017) propose a random walker approach driven by a 3D CNN to segment the esophagus. First, a soft probability map is generated by the CNN. Then, an ACM is fitted to the CNN soft probability map to get a first estimation of the oesophagus location. The outputs of the CNN and ACM are then used in conjunction with a probability model based on Hounsfield Units (HU) to drive the random walker. Training and evaluation were done on 20 clinical and 30 publicly available datasets. Results were assessed regarding spatial overlap and shape similarity, returning a mean Dice coefficient of  $0.76 \pm 0.11$ , an average symmetric surface distance (ASD) of  $1.36 \pm 0.90$  mm, and an average Hausdorff distance of  $11.68 \pm 6.80$ .

Huang et al. (2017) present an automatic osteosarcoma segmentation method, based on FCNs with multiple supervised side output layers. Supervised side output layers were added to three FCNs layers in order to guide the multi-scale feature learning as a contracting structure, and multiple feature channels were used in the up-sampling portion to capture more context information. The results of all the side outputs were fused to determine the final boundaries of the tumours. Quantitative comparison with 405 osteosarcoma manual segmentation results showed an average dice similarity coefficient, average sensitivity, average Hammoude distance and F1-score of 87.80%, 86.88%, 19.81% and 0.908, respectively.

Ibragimov et al. (2017) propose a framework for segmentation of the portal veins (PVs) for liver radiotherapy planning. The authors apply CNNs to learn the consistent appearance patterns of the PVs and then enhance the PVs in previously unseen images. MRFs were further used to smooth the results of the enhancement of the CNN enhancement and remove isolated mis-segmented regions. Finally, CNN-MRF-based enhancement was augmented with PVs centerline detection that relied on PVs anatomical properties such as tubularity and branch composition. The framework was validated on a clinical database with 72 images of patients scheduled for liver stereotactic body radiation therapy. The obtained accuracy of the segmentation was 0.83 and 1.08 mm in terms of the median Dice coefficient and ASD, respectively.

The same authors, in Ibragimov and Xing (2017) use CNNs to study consistent intensity patterns of organ at risks (OARs). For CNNs training, a representative number of positive intensity patches around voxels that belong to the OAR of interest, and negative intensity patches around voxels that belong to the surrounding structures are considered. The trained network is then applied to classify voxels in a ROI in the test image where the corresponding OAR is expected to be located. The obtained classification results are smoothed by using MRFs. Finally, the largest connected component of the smoothed voxels classified as the OAR by the CNN is extracted and dilate-erode operations are performed in order to remove cavities of the component. Performance was validated on the segmentation of spinal cord, mandible, parotid glands, submandibular glands, larynx, pharynx, eye globes, optic nerves, and optic chiasm using 50 CT images, by training one CNNs for each case. The obtained segmentation results varied from 37.4% Dice coefficient for chiasm to 89.5% for mandible.

Men et al. (2017) develop a deep dilated CNN-based method for segmentation of CNNs and OARs. The method is an end-to-end architecture and employed a multiple-scale convolutional architecture to extract multiple-scale context features in the early layers. In addition, the receptive fields of dilated convolutions are enlarged at the end of the network to capture complementary context features. The fully connected layers are replaced with fully convolutional layers to achieve pixel-wise segmentation. Data from 278 patients with rectal cancer

was used for evaluation. The proposed method outperformed the U-Net for all segmentations, with an average Dice value of 3.8% higher than that of U-Net. The authors claim the method to be invariant to the body size, body shape, and age of the patients.

Sun et al. (2017a) are interested in the segmentation of liver tumours. They design a multi-channel FCNs from multiphase contrast-enhanced CTs images. Because each phase of contrast-enhanced data provides distinct information on pathological features, one network is trained for each phase of the CTs images and their high-layer features are then fused. The proposed approach was validated on CT images taken from two databases: 3Dircadb and JDRD. In the case of 3Dircadb, using the FCNs, the mean ratios of the VOE, relative volume difference (RVD), ASD, root mean square symmetric surface distance (RMSD) and maximum symmetric surface distance (MSSD) were  $15.6 \pm 4.3\%$ ,  $5.8 \pm 3.5\%$ ,  $2.0 \pm 0.9\%$ ,  $2.9 \pm 1.5$  mm,  $7.1 \pm 6.2$  mm, respectively. For JDRD, the mean ratios were  $8.1 \pm 4.5\%$ ,  $1.7 \pm 1.0\%$ ,  $1.5 \pm 0.7\%$ ,  $2.0 \pm 1.2$  mm,  $5.2 \pm 6.4$  mm, respectively.

Ait Skourt et al. (2018) propose to use the U-net architecture for lung segmentation. The architecture consists of a contracting path to extract high-level information and a symmetric expanding path that recovers the information needed. This network can be trained end-to-end from very few images and outperforms many methods. Experimental results on LIDC-IDRI database show a Dice coefficient index of 0.9502.

Jackson et al. (2018) design an image segmentation tool based on 3D CNNs to detect right and left kidney contours on non-contrast CTs. The model was trained based on 89 manually contoured cases and tested on a cohort of patients receiving therapy with 177Lu-prostate-specific membrane antigen-617 for metastatic prostate cancer. The neural network segmentation was able to identify right and left kidneys in all patients, returning contours for a selected study in approximately 90 s. Mean Dice score was 0.91 and 0.86 for right and left kidneys, respectively. Poor performance was observed in three patients with cystic kidneys of which only few were included in the training data. No significant difference in mean radiation absorbed dose was observed between the manual and automated algorithms.

Lustberg et al. (2018) investigate whether using software-generated contouring will save time if used as a starting point for manual OAR contouring for lung cancer patients. Twenty CT scans of stage I–III NSCLC patients were used to compare user adjusted contours after an atlas-based and CNN contour, against manual delineation. The CNN was a prototype of deep learning contouring software (“Mirada DLC Expert” prototype, Mirada Medical Ltd., Oxford, United Kingdom) trained on 450 lung patients. The lungs, oesophagus, spinal cord, heart and mediastinum were contoured for this study. With a median time of 20 min for manual contouring, the total median time saved was 7.8 min when using atlas-based contouring and 10 min for CNN contouring. CNN adjustment times were significantly lower than manual contouring time for all OARs.

The aim of the study (Öman et al. 2019) was to investigate the feasibility of ischemic stroke detection from CT angiography using 3D CNNs. Data from 60 patients with a suspected acute ischemic stroke of the middle cerebral artery were randomly selected for this study. Cerebral hemispheric comparison CT angiography and non-contrast CT were studied as additional input features. Authors claim that all ischemic lesions in the testing data were correctly lateralised, and a high correspondence to manual segmentations was achieved. Patients with a diagnosed stroke had clinically relevant regions labelled infarcted with a 0.93 sensitivity and 0.82 specificity. The highest achieved voxel-wise AUC was 0.93, and the highest Dice similarity coefficient was 0.61. When cerebral hemispheric comparison was used as an input feature, the algorithm performance improved. Only a slight effect was seen when non-contrast CT was included.

Xu et al. (2019) split CT slices into image patches, and apply K-Means (KM) with two categories twice using the mean and minimum intensity of image patch, respectively. A cross-shaped verification, a volume intersection, a connected component analysis and a patch expansion follow to generate a train dataset. Using the generated dataset, CNNs are trained. The CNN architecture consists of one convolutional layer with six kernels, followed by one maximum pooling layer and two fully connected layers. A validation experiment is then conducted using a dataset of 201 subjects (4.62 billion patches) with lung cancer or chronic obstructive pulmonary disease. The CNN model achieves an average F-score of 0.9917 and an AUCs up to 0.9991 for classification of lung parenchyma and non-lung-parenchyma. The overlap ratio reaches 0.96.

### Multi-class classification

The previously discussed works applied to segmentation tasks were developed to segment one particular anatomical structure. The next works were used for multi-class classification to create a prediction multi-class map of each voxel to an anatomical label.

Özsavaş et al. (2014) segment the lungs, trachea/main bronchi, and spinal canal of patients with lung cancer to use for radiation treatment planning. For this purpose, an algorithm for inclusion of excluded pathological areas into the segmented lungs and a modified version of the fuzzy segmentation by morphological reconstruction for spinal canal segmentation were developed. The obtained average volume overlap ratio values range between  $94.30 \pm 3.93\%$  and  $99.11 \pm 0.26\%$  on two different datasets, 10 patients with lung cancer and 10 thoracic CT scans from LIDC. Furthermore, the authors obtained ASD values between the ranges of  $0.28 \pm 0.21$  mm and  $0.89 \pm 0.32$  mm by using the same datasets.

Yuan (2017) participated in the MICCAI 2017 Liver Tumour Segmentation Challenge (LiTS) by developing a hierarchical framework based on fully convolutional-deconvolutional neural network (CDNN). A simple CDNN model is firstly trained to provide a quick but coarse segmentation of the liver on the entire CT volume. Then, another CDNN is applied to the liver region for fine liver segmentation. At last, the segmented liver region, which is enhanced by histogram equalisation, is employed as an additional input to the third CDNN for tumour segmentation. IOU is used as loss function when training CDNN models to eliminate the need of sample re-weighting. The framework is trained using the 130 challenge training cases provided by LiTS. The evaluation on the 70 challenge testing cases resulted in a mean Dice similarity coefficient of 0.963 for liver segmentation and 0.657 for tumour segmentation, and a root mean square error (RMSE) of 0.017 for tumour burden estimation.

Zhou et al. (2017a) created a multi anatomical structure segmentation method in 3D CT images employing a voxel-wise multiple-class DL scheme. Their method made use of FCNs that consist of convolution and deconvolution performed on 2D sections from 3D CT images for pixel-wise annotation. By majority voting from multiple crossed 2D sections, the labels were assigned to the 3D segmentation, stacking the 2D labelled results back into 3D. The database consisted of 240 3D CT scans using the simultaneous segmentation of 19 anatomical structures in the human torso divided into 95% for training and 5% for testing. Their own network was trained but no convergence was observed. So, they decided to use VGG16 transfer learning demonstrating the effectiveness of transferring features learned from large-scale datasets to the medical context. They achieved 88.1% and 87.9% of ACC in the train and test, respectively, compared with expert annotations. Their approach exhibited flexibility in segmenting anatomical structures with a wide range of shapes, sizes, appearances, and many different locations through the human body, and the capacity of handling 2D and 3D CT images.

Chmelik et al. (2018) addresses the segmentation and classification of lytic and sclerotic metastatic lesions by using spinal CT obtained from highly pathologically affected cases. The problem is defined as a lesion segmentation based on simultaneous classification of individual voxels to three categories (healthy, lytic, or sclerotic). An individual CNN architecture and pre-processing steps, dependent on a patient data and a scan protocol, are developed. A medial axis transform (MAT) post-processing for shape simplification of segmented lesion candidates with RF based meta-analysis is proposed. The dataset used contained 31 patient cases involving 626 vertebrae (whole-spine CTs were used in 8 cases). The authors claim that the approach achieves a high flexibility—can be applied to variously affected patient data acquired with different CT acquisition parameters. Moreover, the algorithm enables detection, segmentation and classification of small lesions greater than  $1.4 \text{ mm}^3$  and works also with cervical vertebrae.

Gibson et al. (2018) present a registration-free segmentation algorithm for eight organs that are relevant for navigation in endoscopic pancreatic and biliary procedures, including the pancreas, the gastrointestinal tract (oesophagus, stomach, and duodenum) and surrounding organs (liver, spleen, left kidney, and gallbladder). The proposed DenseVNet segmentation network enables high-resolution activation maps through memory-efficient dropout and feature reuse. A multi-centre dataset with 90 subjects is used. The proposed method yielded Dice scores of 0.78 for the pancreas, 0.90 for the stomach, and 0.76 for the oesophagus.

Trullo et al. (2019) make use of global localisation information, based on an original distance map that yields not only the localisation of each organ, but also the spatial relationship between them. Instead of segmenting directly the OARs, a localisation map is first generated by minimising a reconstruction error within an adversarial framework. This map, which includes localisation information of all organs, is then used to guide the segmentation task in a fully convolutional setting. Experimental results on CT scans of 60 patients totalling 11,084 slices are shown for automatic segmentation of the oesophagus, heart, trachea and aorta. The authors comment to have issues with the memory requirements and point out that additional research must be done in order to reduce the size of the system.

Zhu et al. (2019) segment multiple OARs depicted in CT images of lung cancer patients. An encoder-decoder U-Net neural network was used. The trained CNN performed the automatic segmentation of CT images for 36 cases of lung cancer. The Dice similarity coefficient, ASD and 95% Hausdorff distance were calculated, with manual segmentation results used as the standard, and were compared with the results obtained through atlas-based segmentation. For the heart, lungs and liver, both the deep CNN-based and atlas-based techniques performed satisfactorily. For the spinal cord and the oesophagus, the two methods had statistically significant differences.

Table 5 presents a summary of the analysed works with CT applications on segmentation of one or several anatomic interests, while Table 6 refers to conference papers only. The type and the number of instances of each dataset are mentioned. The best performance results are also reported.

**Conclusions** To make the representation of a volumetric image easier to analyse, automatic segmentation is an important task in medical images. Under the application of DL techniques concerning CT modality, the segmentation problem has been mostly approached as a full training problem using CNN or FCN (which is a modification of a CNN as described in Sect. 2) and fewer works have approached the problem by transfer learning. By analysing Tables 5 and 6, we can verify that most automatic segmentation methods have not shown sufficiently robust results for clinical use. This is mainly due to poor image quality, different segmentation protocols, and variations among patients (Zhao and Xie 2013). It is not straightforward to

compare the results of the works reviewed herein. They have been applied with different anatomical structures as targets from various public and private datasets. Each anatomical structure requires levels of performance depending on the purpose. Besides, many evaluation metrics for image segmentation have been employed. To validate the developed methods, the segmentation obtained is compared with manual annotations done by a radiologist who may emphasise diverse aspects (manual annotations are bound to be subjective). Although the performance of the segmentation is context dependent, in general, some conclusions can be drawn:

- Multi-class image segmentation remains a demanding and challenging problem that needs further exploitation as most of the discussed works were developed to segment one particular anatomical structure.
- The development of a non problem specific network architecture or hyperparameters appears to be an ambitious but promising approach.
- The inclusion of lesion cases is expected to make flexible systems that are capable of dealing with heterogeneous structures. However, the problem of having few representative lesion voxels compared to non-lesioned tissue makes the problem highly unbalanced.
- The lack of transfer learning available architectures that conduct 3D operations has repercussions in works currently under development with consequent loss of spatial information.

#### 4.2.2 PET applications

Accurate and reproducible delineation of tumours in PET scans remains a difficult task, despite being crucial for delivering appropriate radiation dose, minimising adverse side-effects of the therapy, and reliable evaluating treatment Czakon et al. (2016). Next, some efforts for the automatising of this task are reviewed.

Czakon et al. (2016) evaluated 3D implementations of Spatial Distance Weighted Fuzzy C-Means (SDWFCM), Dictionary-based Model (DICT) and CNN, and compared them to classical clustering approaches: KM and Gaussian Mixture Models (GMM). The authors tested several CNN architectures and the best, selected through a 5-fold cross-validation process, was a 3D CNN with 3 convolutional layers followed by RELU activations with ascending number of filters. The training dataset was provided by PETSEG challenge and consisted of 4 scans from 3 clinical samples, 9 scans from 3 phantoms and 6 simulated scans in 2 sets. The best results were obtained using CNN (Dice score mean: 0.86, median: 0.89). The remaining methods performed slightly worse than CNN and similarly to each other.

Blanc-Durand et al. (2018) study the feasibility of using 18F-fluoro-ethyl-tyrosine (18F-FET) PET for lesion detection and segmentation. A volumetric CNN was implemented based on a modified Keras implementation of a U-Net library with 3 layers for the encoding and decoding paths. Thirty-seven patients were included in the experiments. All 11 lesions were accurately detected with no false positive, resulting in a sensitivity and a specificity for the detection at the tumour level of 100%. After 150 epochs, the dice similarity coefficient reached 0.7911. Morphological dilatation and fixed thresholding of the predicted U-Net mask improved the dice to 0.8231 (+ 4.1%). At the voxel level, this segmentation led to a 0.88 sensitivity, a 0.99 specificity, a 0.78 positive predictive value (PPV) and a 0.99 positive predictive value (NPV).

Huang et al. (2018) propose a method for head and neck cancer (HNC) gross tumour volume (GTV) contouring on PET/CT. A CNN was trained based on the two-dimensional PET/CT images. The Dice similarity coefficient on a database of 22 newly diagnosed HNC

**Table 7** Main characteristics of the studies reviewed in Sect. 4.2.2: PET segmentation

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Czakon et al. (2016)	Tumours	PETSEG challenge	Dice score mean, median	0.86, 0.89
Blanc-Durand et al. (2018)	Brain	18F-FET PET of 37 patients	Dice, sensitivity, specificity, PPV, NPV	0.8231, 0.88, 0.99, 0.78, 0.99
Huang et al. (2018)	HNC	22 patients	Dice, sensitivity, precision	0.785, 0.764, 0.789
Smith et al. (2018)	Brain	Micro-PET/CT of 28 mice	Dice	0.902
Zhao et al. (2019)	NPC	30 PET/CT scans	Dice, PPV, sensitivity	87.47, 93.56, 82.93

patients was in the range [0.482,0.868]. A high correlation was found between automatic and manual contours, resulting in a median volume difference of 10.9%. The median values of Dice, sensitivity, and precision were 0.785, 0.764, and 0.789, respectively.

Smith et al. (2018) use a pre-trained VGG16 architecture consisting of 16 trainable layers organised in blocks. A total of 28 mice were used for the study ( $0.25 \times 0.25 \times 0.25$ mm). The central sagittal 2D slice and one slice either side (r-l) of the brain were extracted for each animal. Image augmentation using rotations and skews resulted in about 5k images to fine tune the network weights. The mean Dice value found for the test mice is 0.902. The authors claim these results to be comparable to an automated machine-learning based state of the art approach when tested on phantom data and well within the intra-operator variability of manual segmentation.

Zhao et al. (2019) present a method that uses FCN with auxiliary paths to achieve automatic segmentation of Nasopharyngeal carcinoma (NPC) on PET/CT images. The deep supervision introduced by auxiliary paths can explicitly guide the training of lower layers, thus enabling these layers to learn more representative features and improve the discriminative capability of the model. Results of threefold cross-validation return a mean Dice score of 87.47%.

Table 7 presents a resume of the analysed works with PET applications on segmentation of tumours. The type and the number of instances of each dataset are mentioned. The best performance results are also reported.

**Conclusions** In spite of the challenge proposed in MICCAI (Hatt et al. 2018), making a benchmark dataset of 176 images combining simulated, phantom and clinical images available to the community, there has been not a big advance in this field. Among the existing research, Dice similarity coefficient seems to have emerged as the accepted evaluation metric in this line of work.

### 4.2.3 CT and PET applications

Several authors have realised that PET and CT are complementary imaging modalities, one showing functional and the other anatomical information. The idea is thus to use this complementary information to improve segmentation performance.

Zhao et al. (2018a) develop a FCN, capable of taking into account of both PET and CT information simultaneously for tumour segmentation. The network starts with a multi-task

**Table 8** Main characteristics of the studies reviewed in Sect. 4.2.3: CT and PET segmentation

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Zhao et al. (2018a)	Lung	PET/CT of 84 patients	Dice, classification error, volume error	0.85, 0.33, 0.15
Zhong et al. (2018)	Lung	32 PET-CT scans	Dice	0.869 (CT), 0.760 (PET)
Lian et al. (2018)	NSCLC	15 PET/CT	Dice, Hausdorff	0.87, 2.36
Lian et al. (2019)	NSCLC	21 PET/CT	Dice, sensitivity, PPV, Hausdorff, mean absolute surface distance	0.87, 0.86, 0.88, 2.42, 0.19

training module, in which two parallel sub-segmentation architectures constructed using CNNs were designed to extract feature maps from PET and CT respectively. A feature fusion module was subsequently designed based on cascaded convolutional blocks, which re-extracted features from PET/CT feature maps using a weighted cross entropy minimisation strategy. The tumour mask was obtained as the output at the end of the network using a softmax function. The effectiveness of the proposed method was validated on a PET/CT dataset of 84 patients with lung cancer.

Zhong et al. (2018) propose an approach for the segmentation of lung tumours that combines a FCN based semantic segmentation framework (3D-UNet) and the graph cut based co-segmentation model. First, two separate deep UNets are trained on PET and CT, separately, to learn high level discriminative features to generate tumour/non-tumour masks and probability maps for PET and CT images. Then, the two probability maps on PET and CT are further simultaneously employed in a graph cut based co-segmentation model to produce the final tumour segmentation results. Comparative experiments are made on 32 PET-CT scans of lung cancer patients.

Lian et al. (2018) propose a 3D method based on unsupervised learning to jointly delineate tumour contours in PET/CT images. The theory of belief functions is adopted to model the uncertain and imprecise image information, and to fuse them in a stable way. To ensure reliable clustering in each modality, an adaptive distance metric to quantify distortions is proposed, and the spatial information is taken into account. A novel context term is designed to encourage consistent segmentation between the two modalities. In addition, during the iterative process of unsupervised learning, a specific fusion strategy is applied to further adjust results for the two distinct modalities. The method is evaluated in 15 PET/CT images for NSCLC patients. The used database is latter augmented in Lian et al. (2019).

Quantitative results of the above summarised strategies are shown in Table 8.

**Conclusions** On one hand, multimodal segmentation added an obstacle: the fusion of information obtained with CT with the information obtained with PET. On the other hand, it takes advantage of the combination of PET and CT in providing discriminative features for the characterisation of the disease without manually extracting handcrafted features. The analysed studies validate that DL methods can leverage multimodal information. By analysing Table 8, we can conclude that increasing the amount of data seems to be an unequivocal step to further validate the obtained results. There is a need to continue to explore this context with the development of more works in this regard.

## 4.3 Classification

Classification of medical imaging can be separated into two major groups: exam or image classification and lesion classification (Litjens et al. 2017). Exam classification has the purpose of discovering the presence of patterns which can be disease indicators. Commonly, each exam is an input sample and the output is a number that represents the presence or absence of the disease. Rather than whole-body information, lesion classification uses previously segmented body parts (Domingues and Cardoso 2014).

Some works relying on DL techniques with the purpose of classification are revised next for CT and PET. The main characteristics of the analysed studies are summarised in Tables 9 and 10 for CT applications, Table 11 for PET applications and Table 12 for applications that use both CT and PET.

### 4.3.1 CT applications

As seen in Sect. 3, this line of work is the one that has deserved most attention. In order to make this section readable, journal works are described and then summarised in Table 9, whereas conference works are only present in Table 10. Moreover, due to the high prevalence of works related to the lungs, this section is organised into two main parts, one dedicated to Lungs and the other to other organs and anatomical structures.

#### Lungs

Most studies using CT images for CAD are related to lung diseases, particularly lung nodules. Lungs are well defined as dark regions due to their particularly high air percentage and low density. Lung nodules are highlighted in CT images, as they are dense regions that contrast with the lung background.

Chen et al. (2015) developed a system for classification of the malignant or benign nature of nodules with DL techniques, a DBN and a CNN. The method was tested on 2D ROIs extracted from 1010 scans (LIDC-IDRI dataset). All DBNs were pre-trained in an unsupervised manner using RBMs. To evaluate the effectiveness of the DL techniques, two feature-computing methods were performed. The first method relied on scale invariant feature transform (SIFT) and local binary patterns (LBP) features and the kNN was utilised as a classifier. The second method corresponds to a fractal analysis to compute coefficients that were used with the SVM classifier. The best results were obtained with the DBN framework with 73.4% of sensitivity and 82.2% of specificity. They concluded that both DBN and CNN (73.3% of sensitivity and 78.7% of specificity) implementations outperformed the feature-computing methods (75.6% of sensitivity and 66.8% of specificity for SIFT and 50.2% of sensitivity and 57.2% of specificity for LBP).

Cheng et al. (2016) explored the benefits SDAE, a denoising version of SAE. 700 malignant and 700 benign nodules were considered as the experimental dataset for data balanced from the LIDC dataset. The training was divided into two main steps: the pre-training and supervised training steps. At the pre-training step, the ROIs were fed into the network and then refined with the supervised training by the addition of three neurons. The softmax classification performed the final identification result. The performance was evaluated between each pair of the comparing algorithms (SDAE, CURVE, RANK, and MORPH) over all 100 folds. The results showed an increase of performance induced by the SDAE-based CAD algorithm over the conventional systems for the nodule differentiation, with a 0.1 to 0.15 of ACC values higher than CURVE, RANK and MORPH methods. This work was also applied to breast ultrasound lesion.

**Table 9** Main characteristics of the studies reviewed in Sect. 4.3.1: CT classification

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Ciampi et al. (2015)	PFNs	1729 non-contrast-enhanced CT volumes (NELSON database)	AUC	0.868
Chen et al. (2015)	Lung nodules	LIDC-IDRI	Sensitivity, specificity	73.4%, 82.2%
Cheng et al. (2016)	Lung nodules	1400 thoracic CT volumes (LIDC-IDRI)	ACC	0.15
Paul et al. (2016)	NSCLC	81 chest CTs	ACC, AUC	90%, 0.935
Roth et al. (2016)	Sclerotic metastases, LN, Colonic polyp	59 patients (spine imaging), 176 patients (LN), and 1186 patients (colonic polyp)	Sensitivity improvement	57–70% (sclerotic metastases), 43–77% (LNs), and 58–75% (colonic polyps)
Setio et al. (2016)	Lung nodules	888 + 55 thoracic CT scans (LIDC-IDRI + ANODE09)	Sensitivity, FPs per scan, CPM	90.1%, 4, 0.637
van Tulder and de Bruijne (2016)	Lung tissue	73 + 40 lung CT volumes (ILD database + Danish Lung Cancer Screening Trial (DLCST) database)	ACC	Around 70%
Yang et al. (2016)	Lung nodules	1280 thoracic CT volumes (LIDC-IDRI)	Classification error	0.0022–0.48
Ali et al. (2017)	None	16,863 cases with different levels of saturation (heavy, light, and normal)	ACC, error rate	99.92, 0.08
Cha et al. (2017a)	Bladder	123 patients CT scans	AUC	0.69–0.77
Gao et al. (2017)	Brain (AD, lesion and normal ageing)	285 CT brain images	ACC	87.6%
Miki et al. (2017)	Teeth	52 CTs	ACC	88.8%
Shen et al. (2017)	Lung	LIDC-IDRI	ACC, AUC, sensitivity, specificity	87.14, 0.93, 0.77, 0.93
Song et al. (2017)	Lung nodules	4581 thoracic CT volumes (LIDC-IDRI)	ACC, Sensitivity, Specificity, AUC	84.15%, 83.96%, 84.32%, 0.916
Bitbault et al. (2018)	Rectum	95 patients	ACC	80%
de Carvalho Filho et al. (2018)	Lung	50,580 (14,184 malignant and 36,396 benign) nodules (LIDC-IDRI)	ACC, sensitivity, specificity, ROC	92.63%, 90.7%, 93.47%, 0.934

Table 9 continued

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Gao et al. (2018)	ILD	2084 CTs	ACC	87.9%
Jung et al. (2018)	Lung	LUNA16	CPM	0.910
Lee et al. (2018)	Kidney	80 abdominal CTs	ACC	$76.6 \pm 1.4\%$
Liu (2018)	Lung	326 nodules (NLST+ELCAP)	AUC	0.780
Liu et al. (2018b)	Lung	LIDC-IDRI and ELCAP	Classification rate	92.3% (LIDC-IDRI) and 90.3% (ELCAP)
Masood et al. (2018)	Lung	LIDC-IDRI, RIDER, LungCT-Diagnosis, LUNA16 Dataset	ACC	84.58%
Nardelli et al. (2018)	Lung	Right lung of 18 clinical cases with non-contrast chest CTs	ACC	94%
Wang (2018)	Lung	LUNGx	Sensitivity, specificity, AUC	0.896, 8.78, 0.920
Xie et al. (2018)	Lung	LIDC-IDRI	AUC	81.24–96.65%
Yuan et al. (2018)	Lung	LIDC-IDRI and ELCAP	Classification rate	93.1% and 93.9%
Zhao et al. (2018b)	Lung	743 image nodule samples from 1018 scans (LIDC-IDRI)	ACCs, AUCs	0.822, 0.877
Zreik et al. (2018)	Heart	Coronary CT angiography scans of 163 patients	ACCs	0.77 (coronary plaque) and 0.80 (stenosis)
Das et al. (2019)	Liver	225 images	ACC, Jaccard	99.38%, 98.18%
Onishi et al. (2019)	Lung	60 cases	ACC	81.7
Lakshmanaprabu et al. (2019)	Lung	50 LCTs	Sensitivity, specificity, ACC	96.2%, 94.2%, 94.56%
Shaish et al. (2019)	Lymph node	2017 patients	AUC	0.85
Shen et al. (2019)	Lung	LIDC-IDRI	AUC, ACC, sensitivity, specificity	0.856, 0.842, 0.705, 0.889
Xie et al. (2019b)	Lung	LIDC-IDRI	ACC, AUC	91.60%, 95.70%

**Table 10** Main characteristics of the conference works on CT classification

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Da et al. (2015)	Brain	Normal and abnormal CTs	ACC	83.95
Kumar et al. (2015)	Lung	4303 instances containing 4323 nodules (LIDC-IDRI)	ACC, sensitivity, FP/patient	75.01%, 83.35%, 0.39
Roth et al. (2015b)	Neck, lungs, liver, pelvis, legs	1675 axial CT slices	Classification error, AUC	5.9%, 0.998
Shen et al. (2015b)	Lung nodules	1375 thoracic CTs (LIDC-IDRI)	ACC	86.84 %
Gao and Hui (2016)	Brain	282 subjects	ACC	88.81% (AD), 76.7% (lesion) and 95% (normal)
Hong and Sheikh (2016)	Abdominal aortic aneurysm (AAA) region	160 CT slices	Classification error	10–20%
Kim et al. (2016)	Lung	3498 CTs	ACC, Sensitivity, AUC	95.5%, 94.4%, 0.987
Shen et al. (2016)	Lung	2272 lung nodules (LIDC-IDRI)	ACC, AUC	70.69%, 0.66
Hussein et al. (2017)	Lung	1018 scans (LIDC-IDRI)	ACC, mean absolute error (MAE)	91.26%, 0.4593
Nardelli et al. (2017)	Lung	20 clinical cases	ACC	87%
Šprem et al. (2017)	Lung	125 chest CTs	ACC, sensitivity, specificity	85%, 86%, 84%
Antonio et al. (2018)	Spine	15 CTs (xVertSeg Challenge)	ACC	93.29%
Bhatt et al. (2018)	Lung	6306 CT images (LIDC-IDRI)	ACC, sensitivity, specificity	84.13%, 91.69%, 73.16%
Hashimoto et al. (2018)	Lung	High-resolution CTs of 172 patients	ACC	93.3%
Jawali et al. (2018)	Brain	40,367 head CTs	AUCs	0.87
Lei et al. (2018)	Brain	40,367 head CTs	AUCs	0.87
Lyu and Ling (2018)	Lung	1018 cases from 1010 patients (LIDC-IDRI)	ACC	84.81%
Peng et al. (2018)	Lung	91 high-resolution CTs	ACC	92.68%
Sathyan and Pamicker (2018)	Lung	LIDC-IDRI	ACC	98%
Shaffie et al. (2018)	Lung	727 nodules from 467 patients (LIDC-IDRI)	ACC	92.20%
Zahedinasab and Mohseni (2018)	Heart	1729 CTs	ACC	0.807
Chmelik et al. (2019)	Spine	17 cases	Sensitivity	92% at 3 FPs/lesion

Yang et al. (2016) examined the influence of the application of modifications on the original dataset as well as the role of data augmentation with a CNN. Best results were obtained with a balanced dataset of 81,000 images, resulting from the application of transformations such as rotation and scaling to the original dataset composed of 1280 images. Based on their research, they proved that data augmentation plays an integral role in the performance even when simple transformations are used. Also, artificial geometric nodules showed to be ineffective in capturing the important features of real nodules, which reinforces the need for artificial but realistic data.

Shen et al. (2017) investigated the problem of nodule malignancy suspiciousness (the likelihood of nodule malignancy) classification. They presented a Multi-crop CNN (MC-CNN) to automatically extract nodule salient information by employing a multi-crop pooling strategy that crops different regions from convolutional feature maps and then applies max-pooling different times. Experimental results show that the method is capable of characterising nodule semantic attributes (subtlety and margin) and nodule diameter, which are potentially helpful in modelling nodule malignancy.

Song et al. (2017) compared three types of ANNs: CNN, DNN, and SAE. Those networks were evaluated on the LIDC-IDRI database. A total of 4581 images of nodules were used for training with 2265 benign pulmonary nodules samples and 2311 malignant pulmonary nodules samples. The CNN had seven layers, the DNN was composed of six layers and the SAE had five layers. The CNN architecture produced the best results among the three networks, with 84.15%, 83.96%, 84.32% of ACC, sensitivity and specificity, respectively, against 82.37%, 80.66%, 83.9% achieved by DNN and 82.59%, 83.96%, 81.35% achieved by SAE. The AUC of the CNN is 0.916, of the SAE is 0.884, and of the DNN is 0.877. An explanation given was that, unlike other structures, convolution operations could obtain features of different dimensions simultaneously.

Gao et al. (2018) presented an ILD classification approach using CNN with the entire image as input instead of addressing the problem with image patches. The network contained five convolutional layers followed by three fully-connected layers and the final softmax classification layer. The database contained 2084 ROIs labelled from 120 patients split into two subsets for training (100 patients) and testing (20 patients). The output consisted of six categories: normal, emphysema, ground glass, fibrosis, micronodules, and consolidation. Three attenuation ranges were considered and separated into patterns with lower attenuation, higher attenuation and normal lung attenuation, formulating the CNN architecture in three input channels. An overall ACC of 87.9% was obtained.

Liu (2018) investigates four categories of quantitative image biomarkers, including breast density and gynecomastia quantification, bone mineral density, airway dimensions and pulmonary nodule classification. For the classification of pulmonary nodule malignancy, a 3D CNN is trained from scratch. Classifier ensembles of the combinations of the 3D CNN and traditional machine learning models achieve the best performance by taking advantage of the complementary characteristics of the traditional models and the CNN models. The dataset was constructed by combining scans from two large screening studies, National Lung Cancer Screening Trial (NLST) and Early Lung Cancer Action Program (ELCAP), achieving an AUC of 0.780.

Masood et al. (2018) propose a model based on the FCN which is used for classification of pulmonary nodules. The initial classification is done into two classes, nodule (diseased-Malignant or Benign) and non-nodule (normal). The images classified as nodules are further classified into four lung cancer stages. The CNN architecture comprises of 7 convolution with Parametric Rectified Linear Units (PReLU) with  $\alpha = 0.25$ , 7 max-pooling layers, 7 batch normalisation layers and two dense layers with Leaky Rectified Linear Units (LReLU)

with  $\alpha = 0.01$ , a final 1000-dimensional dense layer, deconvolutional layer with a Softmax classifier (large margin softmax loss) on top. In terms of performance, the authors note that it is high even in low-density small-sized pulmonary nodules.

Nardelli et al. (2018) present an approach to classify vessels into arteries and veins. The algorithm follows three main steps: first, a scale-space particles segmentation to isolate vessels; then a 3D CNN to obtain a first classification of vessels; finally, graph-cuts' optimisation to refine the results. The methodology was trained and evaluated on the superior and inferior lobes of the right lung of 18 clinical cases with non-contrast chest CT scans, in comparison with manual classification. The proposed algorithm achieves an overall accuracy of 94%.

Liu et al. (2018b) handle four solid nodule types, well-circumscribed, vascularized, juxta-pleural and pleural-tail. The proposed method is based on multi-view multi-scale CNNs and comprises four main stages. First, the spherical surface centred at nodules is approximated using icosahedra and normalised sampling for CT values are captured on each circular plane at a given maximum radius. Second, intensity analysis is applied based on the sampled values to achieve estimated radius for each nodule. Third, the re-sampling is conducted, followed by a high frequency content measure analysis to decide which planes are more abundant in information. Finally, with approximated radius and sorted circular planes, nodule captures are built at sorted scales and views to pre-train a view independent CNN model and then train a multi-view CNN model with maximum pooling.

Zhao et al. (2018b) propose a CNNs to conquer the challenges of a small-scale medical image database and small size of the nodules. A hybrid CNNs of LeNet and AlexNet is constructed through combining the layer settings of LeNet and the parameter settings of AlexNet. A dataset with 743 image nodule samples is built based on the 1018 scans from LIDC-IDRI to train and evaluate the model. After finely optimising the settings of the CNNs, the estimation ACCs and AUCs can reach 0.822 and 0.877, respectively. The authors note that the accuracy of the CNNs is significantly dependent on the kernel size, learning rate, training batch size, dropout, and weight initialisation. Onishi et al. (2019) use GANs to generate additional images when only small amounts of data are available. Using the proposed method, CTs images of 60 cases with confirmed pathological diagnosis by biopsy are analysed. The benign nodules assessed in this study are difficult for radiologists to differentiate because they cannot be rejected as being malignant. A VOIs centred on the pulmonary nodule is extracted from the CTs images, and further images are created using axial sections and augmented data. A CNNs is trained using nodule images generated by the GANs and then fine-tuned using the actual nodule images to allow the it to distinguish between benign and malignant nodules. This pretraining and fine-tuning process makes it possible to distinguish 66.7% of benign nodules and 93.9% of malignant nodules.

Instead of training a DL model in usual end-to-end way, the following works extracted deep learned representations to be further used as or combined with image descriptors. The responses at each layer, including the input layer, all hidden layers, and the output layer can be used as features and be fed into other classifiers. Briefly, a descriptor outputs a vector whose components describe the content of a contiguous part of an image. Some handcrafted features have considerable discriminative power and they may be complementary to deeply learned features. Intensity, texture and shape descriptors have been used for lung cancer applications. Intensity descriptors are frequently based on statistics computed from grey-level histograms. Texture features are usually based on Grey Level Co-occurrence Matrix (GLCM), wavelet and SIFT features. Regarding shape features, shape index, curvedness and diameter/size have been used (Nogueira et al. 2017).

Ciampi et al. (2015) aimed at classifying peri-fissural nodules (PFNs) in a binary classification task using a pre-trained CNN. The dataset consisted of 1729 CT scans from the

Dutch-Belgian randomized lung cancer screening trial with 568 nodules divided into 65 typical PFNs, 19 atypical PFNs and 484 non-PFNs. The typical and atypical PFNs were coupled to be part of the same PFN label. The dimensionality of 3D pulmonary nodule was reduced through a combination of several 2D views (axial, coronal and sagittal) which were fed into the pre-trained OverFeat architecture to the first classification stage. Next, the given prediction scores were combined for the second classification stage that acts as a combiner to the 3D object. Their experiments included the train of RF and SVM classifiers with OverFeat features from only one of the 2D view of the nodule and with combinations of 2D views. The best results obtained were provided by the features extracted with the Bag of Frequencies descriptor followed by RF classifier achieving an AUC of 0.868.

Paul et al. (2016) aimed at extracting relevant features from NSCLC to predict survivors. The dataset used consisted of chest CT scans distributed into 4 stages: 32 cases in stage 1, 20 in stage 2, 25 in stage 3, and 4 cases in stage 4. Deep features obtained from pre-trained CNN trained on ImageNet (VGG-F, VGG-M, and VGG-S, with 5 convolutional layers followed by 3 fully connected layers) were combined with traditional image features in order to train 4 usual classifiers: naïve Bayes classifier, kNN, decision trees, and RF. Two feature selectors were used and compared, the relief-f and symmetric uncertainty. They obtained 90% of ACC and 0.935 of AUC with the 5 most discriminative features extracted from a VGG-f pre-trained CNN which outperform both the full set of traditional features (77.5% of ACC) and deep features (82.5% of ACC).

de Carvalho Filho et al. (2018) use index basic taxic weights and standardized taxic weights to differentiate the patterns of malignant and benign forms. A CNN is then applied for classification. In the test stage, the methodology is applied to 50,580 (14,184 malignant and 36,396 benign) nodules from LIDC-IDRI database. The proposed method presents achieves ACC of 92.63%, sensitivity of 90.7%, specificity of 93.47%, and ROC of 0.934 for the diagnosis of malignancy and benignity. The authors claim that the results demonstrate a real rate of correct detections using the texture features.

Xie et al. (2018) propose an algorithm for nodule classification that fuses texture, shape and deep model-learned information at the decision level. This algorithm employs a GLCM-based texture descriptor, a Fourier shape descriptor to characterise the heterogeneity of nodules and a CNN to automatically learn the feature representation of nodules on a slice-by-slice basis. It trains an AdaBoosted back propagation ANN using each feature type and fuses the decisions made by three classifiers to differentiate nodules. When the nodules with a composite malignancy rate 3 were discarded, regarded as benign or regarded as malignant, the algorithm achieved an AUC of 96.65%, 94.45% and 81.24% on LIDC-IDRI.

Yuan et al. (2018) use hybrid descriptors consisting of statistical features from multi-view multi-scale CNNs and geometrical features from Fisher vector encodings based on SIFT. First, the nodule radii is approximated based on icosahedron sampling and intensity analysis. Then, high frequency content measure analysis is applied to obtain sampling views with more abundant information. After that, based on re-sampled views, CNNs are trained to extract statistical features and calculate Fisher vector encodings as geometrical features. Finally, hybrid features are achieved by merging statistical and geometrical features based on multiple kernel learning and classify nodule types through a multi-class SVM.

In Lakshmanprabu et al. (2019), the CT scan of lung images is analysed with the assistance of Optimal DNN and Linear Discriminate Analysis (LDA). The deep features are extracted from CT images and then the dimensionality is reduced. The Optimal DNN is applied to CT images and then optimised using Modified Gravitational Search Algorithm for identifying nodules as either malignant or benign. The comparative results show that the proposed classifier gives a sensitivity of 96.2%, specificity of 94.2% and accuracy of 94.56%.

Shen et al. (2019) present an interpretable deep hierarchical semantic CNN to predict whether a given pulmonary nodule is malignant. The network provides two levels of output: (1) low-level radiologist semantic features, and (2) a high-level malignancy prediction score. The low-level semantic outputs quantify the diagnostic features used by radiologists and serve to explain how the model interprets the images in an expert-driven manner. The information from these low-level tasks, along with the representations learned by the convolutional layers, are then combined and used to infer the high-level task of predicting nodule malignancy. This unified architecture is trained by optimising a global loss function including both low- and high-level tasks, thereby learning all the parameters within a joint framework.

The CNN can be supplemented with multiple sources of information as channels. However, when a system is sensitive to small-scale nodules it becomes vulnerable to noise or artefacts that can easily be confused with nodules. For this reason, FP reduction strategies have been developed.

Setio et al. (2016) used multiple streams of 2D CNNs to classify each nodule candidate, obtained by three candidate detectors designed for solid, subsolid, and large solid nodules. To reduce the number of FPs, a multi-view method was proposed, for which sets of 2D patches from differently oriented planes were extracted. 888 scans constituted by 1186 nodules from the LIDC dataset were used to train. The proposed method achieved a sensitivity of 90.1% at 4 FPs per scan. When only considering solid nodules subtype, the proposed approach outperforms other existing CAD approaches with a CPM score of 0.637, demonstrating its generalisation power.

Jung et al. (2018) introduce a method that uses 3D CNNs and an ensemble method to distinguish nodules between non-nodules. A 3D CNNs with shortcut connections and a 3D CNNs with dense connections are used. The shortcut connections and dense connections successfully alleviate the gradient vanishing problem by allowing the gradient to pass quickly and directly. Connections help deep structured networks to obtain general as well as distinctive features of nodules. A checkpoint ensemble method is used to boost performance. Using LUNA16 challenge dataset, the 3D CNNs with shortcut connections and the 3D CNNs with dense connections using the checkpoint ensemble method achieved the highest CPM score of 0.910. The authors claim that 3D CNNs capture the features of spherical-shaped nodules more precisely than their 2D counterparts.

In his MSc thesis, Wang (2018) design a CAD based on a CNN. The CNN is trained on the LUNGx challenge database to classify benign and malignant pulmonary nodules. The architecture of the proposed ConvNet consists of 3 convolutional layers with maximum pooling operations and rectified linear units (ReLU) activations, followed by 2 denser layers with full-connectivities. The proposed CAD system achieved the sensitivity of 0.896 and specificity of 8.78 at the optimal cut-off point of the receiver operating characteristic curve (ROC) with the area under the curve (AUC) of 0.920.

Xie et al. (2019b) propose a multi-view knowledge-based collaborative (MV-KBC) deep model to separate malignant from benign nodules using limited chest CT data. The model learns 3D nodule characteristics by decomposing a 3D nodule into nine fixed views. For each view, a knowledge-based collaborative (KBC) submodel is constructed, where three types of image patches are designed to fine-tune three pre-trained ResNet-50 networks that characterise the nodules' overall appearance, voxel and shape heterogeneity, respectively. The nine KBC submodels are jointly used to classify the nodules with an adaptive weighting scheme learned during the error back propagation, which enables the MV-KBC model to be trained in an end-to-end manner. The penalty loss function is used for better reduction of the false-negative (FN) rate with a minimal effect on the overall performance.

## Other targets

Beyond the classification of pulmonary tissue, the train of DL architecture is also explored with other imaging targets, as reported in the following works.

Roth et al. (2016) explored the possibility of improving the ACC of CAD systems with CNN. Three pre-existing CAD systems were used for selection of lesion candidates on three different datasets: 59 patients for sclerotic metastases detection, 176 patients for LN detection, and 1186 patients for colonic polyp detection. They explored 2D representation from a ROI and 2.5D representation with the extraction of a VOI. Random observations were explored as a data augmentation approach to avoid overfitting and improve the sensitivity with scale transformations, random translations and rotations. Next, these views were used to train deep CNN. Their results showed that CNN could be used to improve CAD performance with reduction of FPs: CAD sensitivities improved from 57% to 70% for sclerotic metastases, 43% to 77% for LNs and 58% to 75% for colonic polyps.

van Tulder and de Bruijne (2016) used RBMs to learn optimised features for 1) lung tissue classification and 2) airway detection. On the lung tissue dataset, the effects of combining generative and discriminative learning objectives were studied. They concluded that adding label information and discriminative learning to the standard RBM helped to produce more appropriate filters than generative or discriminative learning alone, random filters and two standard filter banks: the Leung–Malik and Schmid filter banks. On the airway dataset, the effects of enlarging the size of the representation with more hidden nodes in the RBM, the effects of the filter size and the number of filters on the classification performance were studied. The effect of the number of filters or the number of hidden nodes showed that discriminative learning was less significant for models with more hidden nodes.

Ali et al. (2017) present a classification method based on unsupervised feature extraction and supervised fine-tuning strategies to classify the saturated and unsaturated regions in case of CT saturation. In other words, if protection system is subjected to a CT saturation, the proposed method will classify the different levels of saturation (heavy, light, and normal). Unsupervised feature extraction is made with AEs and CNNs. The authors claim that the results show that the method yields high classification accuracy and has the advantages of simplicity, easier interpretation, model complexity reduction, and better learning capability.

Cha et al. (2017a) studied the impact of radiomics-based predictive models to discern chemotherapy complete or non-complete responses in bladder cancers using pre and post treatment CT images. For that, three models were investigated, DL-CNN, RF-SL (a radiomics feature based approach), RF-ROI (radiomics features from image patterns). 82 patients were used as training data and 41 patients were collected as test set. The RF-SL performed better than DL-CNN and DL-CNN better than RF-ROI which means that the DL-CNN method was able to better capture the characteristics of ROIs in order to identify responders to treatment. The obtained results were in the range of 0.69 to 0.77.

Gao et al. (2017) inspected the application of an integrative 2D and 3D CNNs for the classification of CT brain images divided into three groups (57 Alzheimer disease (AD), 115 lesions and 113 normal ageing) with the intent of providing information for the early diagnosis of AD. Image slices were used to train the 2D CNN model while in the 3D path, small cubes were applied. The fusion of the two CNNs seven layers networks was made based on the average of softmax scores obtained from both networks that determine a score of normalised class probabilities. The authors, surprisingly, claim that 2D approaches outperform their 3D counterparts. The justification lies in the fact that CT data had low depth resolution leading to worsen the 3D results. The proposed 2D/3D CNN fusion method achieved the best results

with 87.62% of ACC rate, indicating that the combination of 2D and 3D creates a synergy capable of obtaining characteristics obtained in each network.

Miki et al. (2017) aim at automating the dental filing process by using dental X-ray images. ROIs including single teeth are extracted from CT slices and classified into 7 tooth types. Fifty two CT volumes were randomly divided into 42 training and 10 test cases, and the ROIs obtained from the training cases were used for training an AlexNet CNN. For examining the sampling effect, random sampling was performed 3 times, and training and testing were repeated. The average classification accuracy using the augmented training data by image rotation and intensity transformation was 88.8%. The authors highlight that, with this approach there is no need for precise tooth segmentation.

Bibault et al. (2018) predict complete response from pre-treatment features. Radiomic features were extracted from the tumour volume from the treatment planning CT and a DNN was created to predict complete response. The results were compared to LR and SVM. Ninety-five patients were included in the final analysis, of which 22 (23%) had pathologic complete response after chemoradiation. The DNN predicted complete response with an 80% accuracy, which was better LR (69.5%) and SVM (71.58%). These results may help to identify patients who would benefit from a conservative treatment, rather than a radical resection.

Lee et al. (2018) differentiate small renal masses (SRM) into angiomyolipoma without visible fat (AMLwvf) and clear cell renal cell carcinoma (ccRCC) using a concatenation of hand-crafted and deep features, and RF classifier. The 1000–4000-dimensional deep features were extracted from the ImageNet pretrained deep learning model with the SRM image patches. In experiments, the combinations of four deep learning models, AlexNet, VGGNet, GoogleNet, and ResNet, and four input image patches, including original, masked, mass-size, and texture image patches, were compared and analysed. The authors observe that (a) the concatenation of features outperformed using either set of features alone, (b) AlexNet showed generally the best performances among the CNN models, and (c) texture image patches not only achieved competitive performances among the input patches, but also steady performance regardless of CNN models.

Zreik et al. (2018) apply a multi-task recurrent CNN on multi-planar reformatted images of the coronary arteries on coronary CT angiography scans of 163 patients. First, a 3D CNN is utilised to extract features along the coronary artery. Subsequently, the extracted features are aggregated by a RNN that performs two simultaneous multi-class classification tasks. In the first task, the network detects and characterises the type of the coronary artery plaque (no plaque, non-calcified, mixed, calcified). In the second task, the network detects and determines the anatomical significance of the coronary artery stenosis (no stenosis, non-significant i.e. < 50% luminal narrowing, significant i.e. > 50% luminal narrowing). For detection and classification of coronary plaque, the method achieved an accuracy of 0.77. For detection and classification of stenosis, the method achieved an accuracy of 0.80.

Das et al. (2019) propose a watershed Gaussian based DL technique for effective delineate the cancer lesion in CT images of the liver. Initially, the liver is separated using marker controlled watershed segmentation process and then the cancer affected is segmented using GMM. After tumour segmentation, various texture features are extracted from the segmented region. These segmented features were fed to a DL classifier for automated classification of three types of liver cancer i.e. hemangioma, hepatocellular carcinoma and metastatic carcinoma. A classification accuracy of 99.38%, Jaccard index of 98.18%, were achieved in a 225 images database.

Shaish et al. (2019) try to predict the maximum standardised uptake value (SUV) of lymph nodes. Consecutive initial staging PET/CT scans from 2017 patients with pathologically

proven malignancy were collected. Two blinded radiologists selected one to 10 lymph nodes from the unenhanced CT portion of each PET/CT examination. Lymph nodes were cropped and used with the primary tumour histology type as input to a 3D CNN. An SUV of 2.5 or greater was defined as FDG avid. The predicted SUV was associated with the real SUV and with FDG avidity, with an AUC value of 0.85.

Table 9 recapitulates the main characteristics of the analysed works and Table 10 of the conference works with CT applications on classification task. The type and the number of instances that are included in the datasets are mentioned as well as the best performance results.

**Conclusions** In the classification context, a variety of DL techniques, both supervised and unsupervised, have been used, namely CNN, SAE, SDAE, DNN, DBN, and RBM, as full training classifier, deep feature extractors, and transfer learning. The works focused on feature extraction can be divided into: works that used DL techniques as feature learning to be further fed into a traditional classifier (with and without the combination of features obtained by traditional descriptors), works that extracted features from CNNs pre-trained in non-medical images, and works that used deep architectures to classify the training instances fed with features obtained by traditional descriptors. These studies are limited in two aspects:

- The features were often extracted with 2D approaches, lacking features based on volumetric information that may be significant to pathology stage and treatment assessment.
- The models were built taking into consideration a single diagnostic data source, without considering the relationship within multiple sources of data that are commonly used in the clinical practice.

Observing Tables 9 and 10, it is clear that most studies using CT data for classification systems are applied to lung diseases, particularly lung nodules. The reason is two-fold: lung nodules are well identified in CT images and several public datasets are available. A common problem in these studies is the FP rate, due to the fact that some anatomical locations naturally exhibit high density and are mistaken for tumours.

#### 4.3.2 PET applications

FDG-PET images are not only useful for the analysis of pathologies characterised by regions of abnormally high uptake, as tumours but also for low metabolism regions as AD.

Ypsilantis et al. (2015) investigate the possibility of creating a system with the capability to predict, from a scan acquired before the treatment, the patient's reaction to chemotherapy. They compare two main approaches, tested on samples from 107 patients with esophageal cancer. The standardised one was to extract a large number of features with recurrent image descriptors such as GLCM, followed by its application in statistical classifiers: LR, gradient boosting, RFs, and SVMs. On the other hand, in order to use directly the data, they used a CNN, 3S-CNN. The inputs were triplets of adjacent slices of the ROI after having segmented the tumour. In addition, they implemented a simpler architecture 1S-CNN which considers each slice individually. In this case, as they were only focused on a very specific esophageal cancer, they have ignored spatial and anatomical information which does not include tumours that spread throughout the body. Their system is highly dependent on an effective pre segmentation of the tumour. They claim to achieve average 73.4% ACC with the 3S-CNN algorithm.

Cheng and Liu (2017) proposed a system based on the combination of several 2D CNNs and RNNs for AD classification diagnosis from 3D PET images. Their method was evaluated

on PET images from 339 subjects including 93 AD patients, 146 mild cognitive impairment (MCI) and 100 normal controls (NC) from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The PET volume was firstly decomposed into 2D images slices in coronal, sagittal, and axial axis and the resulting slices were assembled into groups each composed of 15 slices. Thus, each group were used as an input of 2D CNN with the purpose of capturing intra-slice attributes. Each deep CNN consists of 5 convolutional layers, 2 max pooling layers, and 2 fully connected layers and 1 softmax layer. To discover inter-slice features, two stacked Bidirectional-Gated Recurrent Units (GRUs) of RNN were implemented with the feature vectors as input, previously produced by 2D CNNs. Then, the network was followed by two fully connected layers and a softmax layer to make the final classification prediction. Their method achieved AUC of 95.28% for classification of AD versus NC and 83.90% for classification of MCI versus NC.

Singh et al. (2017) also worked on the AD diagnosis problem. The developed method consisted of max-pooling and mean-pooling on the 3D PET images to make them 2D and to reduce the dimension of features from  $79 \times 95 \times 79$  to a one-dimensional (1D) vector of size 4050. Additionally, further dimensionality reduction using Probabilistic Principal Component Analysis was performed reducing the number of features from 4050 to the range of 250 to 300 of features. Finally, a five hidden layers MLP performed binary classification. They complete their study with the inclusion of demographic features, age and gender, which improved prediction results. The dataset consisted of 186 cognitively unimpaired (CU) subjects, 336 MCI subjects with 158 Late MCI and 178 Early MCI, and 146 AD patients. The F1-score from the best results for the classification of FDG-PET data with and without demographics additional features was 0.9812 and 0.9430 respectively. Also, max-pooling achieved superior performance in the majority of the binary classification experiments compared to mean-pooled features.

Choi and Jin (2018) aimed to develop a system to predict future MCI patients and select subjects that could convert to AD, using PET images of 139 patients with AD, 171 patients with MCI and 182 normal subjects. The first part corresponded to train a CNN with AD and NC for classification purposes. The architecture used consisted of three convolution layers and one fully connected layer with two types of PET data as two channel data: FDG and AV-45 (a 4D vector as input). Following, the trained network was further used to predict cognitive outcome in MCI subjects, whether they would transform to AD or not. They claimed to achieve 96% of ACC for the differentiation between AD and NC and 84.2% of ACC in predicting MCI conversion.

Liu et al. (2018a) propose a classification framework based on the combination of 2D CNNs and RNNs, which learns the intra-slice and inter-slice features for classification of AD after decomposition of the 3D PET image into a sequence of 2D slices. The 2D CNNs are built to capture the features of image slices while the GRUs of RNNs is cascaded to learn and integrate the inter-slice features for image classification. The method is evaluated on the baseline FDG-PET images acquired from 339 subjects including 93 AD patients, 146 MCI and 100 NC from ADNI database. Experimental results show that the method achieves an AUC of 95.3% for AD versus NC classification and 83.9% for MCI versus NC classification.

Pereira (2018) focus on the evaluation of response to treatment in Hodgkin Lymphoma. The pipeline includes an image registration step to extract the VOIs (Pereira et al. 2018), and data augmentation with simultaneous random rotations on the three axes and with rotation, scaling, shear, and skew transformations. Experiments with several different CNNs architectures were performed, and best results were achieved with a 4 layer network with two convolutional layers followed by two fully connected layers for 750 training samples and a batch size of 5. The results obtained from a real dataset of 163 patients with Hodgkin

Lymphoma from IPO-Porto achieved 0.88 of MAE and 0.99 of average MAE in classifying the exams into the 5 classes Deauville scale.

Schwyzer et al. (2018) evaluated whether machine learning may be helpful for the detection of lung cancer in FDG-PET imaging in the setting of ultralow dose PET scans. A transfer learning approach to perform binary classification of 2D PET image slice was implemented by using a pre-trained deep residual neural network. The database consists of 3936 PET slices including images in which the lung tumour is visually present and image slices of patients with no lung cancer were exported. The AUC was 0.989, 0.983 and 0.970 for standard dose images (PET100%), and reduced dose PET10%, and PET3.3% reconstruction, respectively. A sensitivity of 95.9% and 91.5% and a specificity of 98.1% and 94.2%, at standard dose and ultralow dose PET3.3%, respectively, were also obtained.

Table 11 highlights the main characteristics (anatomic interest, dataset quantification, evaluation metric and best performance result obtained) of the analysed works with classification purpose applied on PET modality alone.

**Conclusions** Under PET modality, full training approaches have been explored for binary classification. The majority of the works decomposed the 3D image into a sequence of 2D image slices, not taking into account volumetric information which may be important for the classification of certain degrees of disease. Analysing Table 11, we can easily verify that there has been a growing interest to FDG-PET for AD and AD risk. The availability of public datasets can be a reason for this tendency. PET information processing is an under-studied topic. We anticipate that these works will inspire more research that the building of DL based systems for FDG-PET data analysis.

### 4.3.3 CT and PET applications

As previously seen, some work using information from both PET and CT, fusing the information somewhere along the proposed pipeline. Fusion can typically be at the feature level, at the classifier level, or at the decision level Guo et al. (2018).

Qiang et al. (2017) proposed a pulmonary nodule diagnosis approach with feature extraction using a deep supervised AE. The dataset consisted of 120 PET/CT scans from patients with lung cancer including 2810 pulmonary nodules. A high-level discriminative feature representation in PET/CT was extracted from SDAE based on extreme learning machine (ELM), SDAE-ELM. Next, two methods were used to fuse the dual-modal features information: decision level fusion where PET and CT images were processed independently using one unique classifier for each, and feature level fusion where one LR is used for the concatenation of both modalities. Their method reached 91.75% of sensitivity at 1.58 FPs per scan with better discriminative results than other DL diagnosis methods like AE, deep AE and RBM.

For classifying mediastinal LN metastasis from FDG PET/CT images, Wang et al. (2017a) compared a DL method with four other mainstream machine learning. The five methods were evaluated using 1397 LNs collected from PET/CT images of 168 patients. The CNN architecture used was AlexNet with five layers and the inputs were axial image patches cropped from the CT and PET SUV images. The authors claim to demonstrate that the performance of CNN is not notably different from the long-established methods (sensitivity = 84%, specificity = 88%, ACC = 86%, AUC = 0.91). Although classical machine learning techniques provided the best results (ACC = 81–85%, AUC = 0.87–0.92), they were optimised and tested with a variety of parameters. Only one CNN architecture was tested without further research on improving the performance.

**Table 11** Main characteristics of the studies reviewed in Sect. 4.3.2: PET classification

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Ypsilantis et al. (2015)	Esophageal cancer	107 FDG-PET	ACC	73.4%
Cheng and Liu (2017)	AD	339 subjects including 93 AD patients, 146 MCI and 100 NC (ADNI dataset)	AUC	95.28% for classification of AD versus NC, 83.90% for classification of MCI versus NC
Singh et al. (2017)	AD	186 CU subjects, 336 MCI subjects with 158 Late MCI and 178 Early MCI, and 146 AD patients	F1	0.9812 with demographics additional features, 0.9430 without demographics additional features
Choi and Jin (2018)	AD	139 patients with AD, 171 patients with MCI and 182 normal subjects (ADNI-2 dataset)	ACC	96% for the differentiation between AD and NC, 84.2% for predicting MCI conversion
Liu et al. (2018a)	AD	93 AD patients, 146 MCI and 100 NC (ADNI database)	AUC	95.3% for AD versus NC classification and 83.9% for MCI versus NC classification
Pereira (2018)	Hodgkin Lymphoma	163 patients from IPO-Porto	MAE, MAE	10.88, 0.99
Schwyzler et al. (2018)	Lung	3936 PET slices	AUC, sensitivity, specificity	0.989, 95.9%, 98.1% (standard dose); 0.970, 91.5%, 94.2% (ultralow dose)

**Table 12** Main characteristics of the studies reviewed in Sect. 4.3.3: CT and PET classification

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Qiang et al. (2017)	Pulmonary nodules	120 PET/CT scans from patients with lung cancer including 2810 pulmonary nodules	Sensitivity, FPs per scan	91.75%, 1.58
Wang et al. (2017a)	LN	1397 LNs collected from PET/CT images of 168 patients	Sensitivity, specificity, ACC, AUC	84%, 88%, 86%, and 0.91

Table 12 presents the characteristics (anatomic interest, dataset, evaluation metric and best performance result obtained) of the examined works with classification purpose integrating PET and CT information.

**Conclusions** As distinctive works with different purposes were analysed in this section, a direct comparison of the results represented in Table 12 is not meaningful. The first one employed a CNN in a full training way for LN metastasis classification. Despite lung cancer diagnosis being mainly done recurring to CT images (as stated on Sect. 4.3), the second work employed SDAE to extract deep features from PET/CT images for pulmonary nodule diagnosis achieving comparable results. However, these studies have a common point. Images from both modalities, PET and CT, were mixed into the network. The fusion is, however, done at either feature or decision level. No attempts of fusion the information at the classifier level have yet been made.

## 4.4 Other applications

The following works describes others contributions of DL techniques to various understudied tasks, namely noise and artefacts reduction, reconstruction, estimating TOF, generating PET images, and radiation adaptation. The main features of the analysed studies are summarised in Tables 13, 14, 15, 16, and 17.

### 4.4.1 CT applications: noise and artefacts reduction

Due to growing concerns about the overall radiation exposure to a patient, CT devices have been designed to lower the dose (Costello et al. 2013). The most common way to lower the radiation dose is to reduce the X-ray flux by decreasing the operating current and shortening the exposure time of an X-ray tube. However, such a technique results in reduced image quality with the cost of the appearance of many artefacts, which may obscure or simulate pathology (Boas and Fleischmann 2012).

Chen et al. (2017a) used a residual encoder-decoder CNN to strengthen the quality of LCT images to normal-dose images. Their method combined AE and CNN consisting of 10 layers, including 5 convolutional and 5 deconvolutional layers. From the encoder-decoder network, a residual learning with shortcut connections was added to supplement the structure lost through multiple layers. The normal dataset included 7015 normal-dose CT images from the National Cancer Imaging Archive (NCIA) and the low-dose images were produced by applying Poisson noise. The proposed network achieved, in the testing dataset: peak-signal-

**Table 13** Main characteristics of the studies reviewed in Sect. 4.4.1: CT noise and artefacts reduction

Authors	Dataset	Evaluation metrics	Results
Chen et al. (2017a)	7015 normal-dose CT images from 165 patients (NCIA)	PSNR, RMSE, SSIM	$43.7871 \pm 2.01$ , $0.0069 \pm 0.0007$ , $0.9754 \pm 0.0086$
Chen et al. (2017b)	7015 normal-dose CT images from 165 patients (NCIA)	PSNR, RMSE, SSIM	$42.1514$ , $0.0080$ , $0.9707$
Kang et al. (2017)	5743 (3642 + 2101) abdominal CT slices (Low Dose CT Grand Challenge)	Visual inspection	Subjective assessment
Wolterink et al. (2017)	non-contrast-enhanced cardiac CT scans of 28 patients	PSNR	40–45
Shan et al. (2018)	normal-dose abdominal CT from 10 patients (Mayo Challenge) and corresponding simulated quarter-dose CT	PSNR, SSIM, perceptual loss, texture matching loss	$30.136$ , $0.905$ , $1.85$ , $67.63$ .
Tang et al. (2018)	DeepLesion dataset, composed of 32,735 PACS CT lesion images	Recall, precision, dice	$0.911$ , $0.940$ , $0.920$
Wang et al. (2018a)	Mayo challenge	PSNR, feature similarity (FSIM), normalized mean squared error (NMSE)	$32.6947$ , $0.8919$ , $0.0417$
Yang et al. (2018)	Mayo challenge	PSNR, SSIM	$22.0168$ , $0.7745$
Yi and Babyn (2018)	CT scans of a piglet at different dose levels, from 300 to 15 mAs	SSIM, noise level	$0.95$ , $8.80$
You et al. (2018)	Mayo challenge (3 scans)	PSNR, SSIM, RMSE	$26.446$ , $0.786$ , $0.047$
Zhang and Yu (2018)	Synthetic data based on clinical CT images (Low Dose CT Grand Challenge)	RMSE, SSIM	$22.8$ , $0.977$
He et al. (2019)	Mayo Challenge	PSNR, NMSE, FSIM	$41.2026$ , $0.0185$ , $0.9442$ (20 mAs); $39.1636$ , $0.0234$ , $0.9332$ (12 mAs); $37.0485$ , $0.0266$ , $0.9214$ (4 mAs)
Lee et al. (2019)	TCIA	SSIM	$0.9$
Maijer et al. (2019)	CT scan of an anthropomorphic head phantom	Deviation	$6$ HU
Wang et al. (2019)	Brain CT	SSIM, PSNR	$0.913$ , $30.31$

**Table 14** Main characteristics of the studies reviewed in Sect. 4.4.2: PET reconstruction

Authors	Dataset	Evaluation metrics	Results
Cui et al. (2017)	Brain phantom and Zubal phantom, real whole-body PET to extract heart data	SNR, Bias, Variance	70.42, 0.0365, 0.0071 for Brain phantom and 68.41, 0.0369, 0.009 for Zubal phantom
Jiao and Ourselin (2017)	Generated data from ADNI database	Visual inspection	Subjective assessment
Spencer and Wang (2017)	Simulated dynamic PET scans for a Zubal head phantom	MSE	1–5
Kim et al. (2018)	High Resolution Research Tomograph data acquired for a schizophrenia dynamic study on 27 control subjects and clinical patients	normalised RMSE, SSIM	0.266, 0.496
Gong et al. (2018)	70-min dynamic PET scan of a human subject	Contrast recovery coefficient versus standard deviation	uptake between 0.7 and 0.8 for SD between 6 and slightly over 6
Gong et al. (2019)	6 lung patient data and 17 brain scans	Contrast recovery coefficient versus standard deviation	uptake between 0.6 and 0.8 for SD between 5 and 2; uptake between 0.65 and 0.75 for SD between 5 and 10

**Table 15** Main characteristics of the studies reviewed in Sect. 4.4.3: PET estimating TOF

Authors	Dataset	Evaluation metrics	Results
Berg and Cherry (2018)	15,000 coincidence waveforms	Improvement of coincidence timing resolution	23%

**Table 16** Main characteristics of the studies reviewed in Sect. 4.4.4: PET generation

Authors	Dataset	Evaluation metrics	Results
Bi et al. (2017a)	50 PET/CT studies volumes from 50 lung cancer patients	MAE, PSNR	4.60, 28.06
Ben-Cohen et al. (2017)	25 CT and PET pairs constrained to the liver region	TPR, FPR	92.3%, 0.25
Ben-Cohen et al. (2019)	60 CT and PET pairs constrained to the liver region	MAE, PSNR	0.72, 30.22

**Table 17** Main characteristics of the studies reviewed in Sect. 4.4.5: PET radiation adaptation

Authors	Anatomic interest	Dataset	Evaluation metrics	Results
Tseng et al. (2017)	NSCLC	114 NSCLC patients	RMSE	0.5 Gy
Xiang et al. (2017)	Not specified	LPET, SPET, MRI for 16 subjects	NMSE, PSNR	24.76, 0.0206
Wang et al. (2018b)	MCI	16 subjects (8 normal and 8 MCI)	NMSE, PSNR	24.5, 0.025

to-noise-ratio (PSNR) =  $43.7871 \pm 2.01$ , RMSE =  $0.0069 \pm 0.0007$ , and structural similarity index (SSIM) =  $0.9754 \pm 0.0086$ .

The same team, in Chen et al. (2017b), proposes a noise reduction method for LCT via deep learning without accessing original projection data. A CNN is used to map LCT images towards their corresponding normal-dose counterparts in a patch-by-patch fashion. Results were assessed using 7015 CT normal-dose images from 165 patients including different parts of the human body, downloaded from NCIA. In terms of quantitative metrics, PSNR, RMSE and SSIM were used, achieving 42.1514, 0.0080, and 0.9707, respectively.

Kang et al. (2017) proposed an algorithm for LCT reconstruction that uses deep CNN. In order to decompose the components of noise, a directional wavelet transform as a contourlet transform was used. Then, the deep networks composed of 24 convolution layers were used to de-noise accepting wavelet coefficients as input. In parallel, for comparison purposes, a baseline CNN was trained whose input and output layer were images. The dataset was generated from the projection data of 2016 CT low-dose Grand Challenge using a filtered backprojection algorithm. The original data was composed of 3642 slices of 3D CT projection data with normal-dose and quarter-dose for the training set and 2101 slices for the testing set.

Their experiments proved that the wavelet-domain CNN was effective in removing complex noise patterns compared to existing model-based iterative reconstruction methods.

Wolterink et al. (2017) propose to train a CNN jointly with an adversarial CNN to estimate routine-dose CT from LCT and hence reduce noise. A generator CNN was trained to transform LCT into routine-dose CT using voxelwise loss minimization. An adversarial discriminator CNN was simultaneously trained to distinguish the output of the generator from routine-dose CT. The performance of this discriminator was used as an adversarial loss for the generator. Experiments were performed using non-contrast-enhanced cardiac CT scans of 28 patients. For each patient, two scans were acquired, at 20% and 100% routine clinical dose. The results showed that training with voxelwise loss resulted in the highest PSNR with respect to reference routine-dose images.

Shan et al. (2018) introduce a conveying path-based convolutional encoder-decoder (CPCE) network in 2D and 3D configurations within the GAN framework for LCT denoising. A novel feature of this approach is that an initial 3D CPCE denoising model can be directly obtained by extending a trained 2D CNN, which is then fine-tuned to incorporate 3D spatial information from adjacent slices. Based on the transfer learning from 2D to 3D, the 3D network converges faster and achieves a better denoising performance when compared with a training from scratch. Experimental data stem from an authorised clinical LCT dataset, which was made for the 2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge. This dataset included normal-dose abdominal CT images that were taken from 10 anonymous patients and the corresponding simulated quarter-dose CT images. Results are as follows: PSNR = 30.136, SSIM = 0.905, perceptual loss = 1.85, and texture matching loss = 67.63.

Tang et al. (2018) focus on a pre-processing method for CT that uses stacked GAN (sGAN) with the goal of improving lesion segmentation. The first GAN reduces the noise in the CT image and the second GAN generates a higher resolution image with enhanced boundaries and high contrast. Tests have been made on whether sGAN can yield improved lesion segmentation using as metrics recall, precision and Dice score. Experimental results on the DeepLesion dataset demonstrate that sGAN is more effective than four other enhancement approaches

Wang et al. (2018a) propose an iterative residual-artefact learning CNN iterative residual-artefact learning CNN (IRLNet) approach for reconstruction. Specifically, the proposed IRLNet estimates the high-frequency details within the noise and then removes them iteratively; after eliminating severe streaks in the LCT, the residual low-frequency details can be processed through the conventional network. Moreover, the proposed IRLNet scheme can be extended for robust handling of quantitative dual energy CT/cerebral perfusion CT imaging, and statistical iterative reconstruction. Data from 2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge was used for evaluation and the experimental results demonstrate that IRLNet outperforms other CNN based approaches in reducing the image noise and streak artefacts efficiently at the same time as preserving edge details.

Yang et al. (2018) introduce a GAN with Wasserstein distance and perceptual similarity. The Wasserstein distance is a key concept of the optimal transform theory, and improves the performance of the GAN. The perceptual loss compares the perceptual features of a denoised output against those of the ground truth in an established feature space, whereas the GAN helps migrate the data noise distribution from strong to weak. Experiments made on patches extracted from the Mayo Challenge dataset demonstrated that using the proposed GAN helps improving image quality and its statistical properties.

Yi and Babyn (2018) train an adversarial network and a sharpness detection network to guide the training process. CT scans of a deceased piglet were obtained at different dose levels, ranging from 300 mAs down to 15 mAs. At each dose level, 850 images of a size  $512 \times 512$

were obtained. The mean standard deviation of 42 smooth rectangular homogeneous regions (size of  $21 \times 21$ ,  $172.27 \text{ mm}^2$ ) was computed as measure of the noise level. Experiments show that the results of the proposed method have small resolution loss and achieve better performance relative to the compared state-of-art methods both quantitatively and visually.

You et al. (2018) propose a 3D noise reduction method, called structurally sensitive multi-scale generative adversarial net, to improve the LCT image quality. Different loss functions for training denoising models are investigated. Experiments using the Mayo dataset show that the method can effectively preserve the structural and textural information in reference to the normal-dose CT images and significantly suppress noise and artefacts. Qualitative visual assessments by three radiologists further demonstrate that the method retrieves more information and outperforms competing methods.

With the purpose of suppressing metal artefacts, Zhang and Yu (2018) exploited a CNN, which merges data from the original and corrected images. The database was generated using 74 metal-free and 15 metal shapes (such as dental fillings, hip prostheses, wires). In total, 100 samples were created. Following, 10,000 patches from the database were extracted to train the CNN composed of 4 convolutional layers according to the following workflow. The metal segmentation and artefact reduction were initially done. Next, another artefact reduction was performed with the trained CNN to replace metal affected projections with the forward projection, followed by the filtered back projection reconstruction. The effectiveness of the proposed method was also validated on clinical data. Experimental results demonstrated that the proposed method was able to distinguish tissue structures from artefacts, obtaining a RMSE of 22.7 of the original and corrected images with respect to the reference images and a SSIM of 0.977 in the numerical simulation study.

He et al. (2019) propose a strategy for model-based iterative reconstruction (MBIR) that addresses prior knowledge design and MBIR parameter selection in one optimisation framework. Specifically, a parameterized plug-and-play alternating direction method of multipliers is proposed for the general penalised weighted least-squares model, and then, the parameterised plug-and-play prior and the related parameters are optimised simultaneously in a single framework using a large number of training data. Experimental results obtained on clinical patient datasets demonstrate that the proposed method can achieve promising gains over existing algorithms for LCT image reconstruction in terms of noise-induced artefact suppression and edge detail preservation.

Lee et al. (2019) used a FCN and a wavelet transform to predict high quality images. To reduce the spatial resolution loss of predicted images, the pooling layer was replaced by a wavelet transform. Three different domains were evaluated—the sinogram domain, the image domain, and the hybrid domain. To train and develop the model, TCIA dataset was used. The authors claim that streak artefacts, which generally occur under sparse sampling conditions, were effectively removed. However, image characteristics of fine structures varied. In the sinogram domain, some image noise remained; in the image domain, a blurring effect occurred. The hybrid domain sparsely sampled reconstruction was able to restore images to a quality similar to fully sampled images.

Maier et al. (2019) present a CNN to derive scatter estimates in real time. Since gold standard scatter estimation approaches rely on MC photon transport codes, the proposed CNN is trained to predict MC scatter estimates based on the acquired projection data. The potential of the proposed strategy is demonstrated using simulations of cone-beam CT head, thorax, and abdomen scans as well as measurements at an experimental table-top cone-beam CT. To evaluate the performance for real data, measurements of an anthropomorphic head phantom were performed. Errors were quantified by a comparison to a slit scan reconstruction.

The achieved deviation is 6 HU. The authors also note that the proposed method is nearly as accurate as the traditional MC simulations but is superior in terms of speed.

Wang et al. (2019) is interested in removing ring artefacts from cone-beam CT, while retaining important texture details. The method starts by transforming ring artefacts in Cartesian coordinates to stripe artefacts in polar coordinates. A GAN loss function is then designed, which includes three parts: unidirectional relative total variation loss, perceptual loss and adversarial loss. Very deep residual networks are used for both the generator and discriminator. As conclusions, the authors state that the method improves the ability to generalisation and fidelity for removing ring artefacts.

Table 13 outlines the main characteristics (anatomic interest, dataset, and best performance results) of the previously discussed works on the CT modality.

**Conclusions** The success of LCT is determined by both the performance of the denoising techniques and the capability of suppressing artefacts. Noise and artefact reduction are important concepts and thus the research interest has been significant in CT dose reduction. The quality of the acquisition makes large differences in how the images can be analysed. Similarly to other applications of DL in CT, CNNs are the most popular technique. Analysing Table 13, evaluation metrics based on the measurement of the similarity between two images have been used, as well as visual inspection which is highly prone to subjectiveness. Most of the datasets used were either artificial or composed of a very small number of scans. It would be valuable to further test these approaches on large datasets used on real clinical use.

#### 4.4.2 PET applications: reconstruction

In opposition to static PET imaging, dynamic PET imaging detects data in a series of frames. The reconstruction has to guarantee temporal resolution. Image reconstruction can be considered as a dense prediction problem, where for each voxel/pixel, a model predicts a value in the reconstructed image from the input data.

Cui et al. (2017) developed a reconstruction system for dynamic PET imaging preserving key features based on deep SAE and maximum likelihood expectation maximisation (MLEM), with noise reduction. The process was divided into two main parts: the training in which series of reconstructed images were the input of a SAE model and the reconstruction part, firstly by MLEM, and then the reconstructed images were used as input to the SAE model with the corresponding parameters obtained on the previous training. Due to the dynamic nature of reconstruction, temporal features were also learned by the single-layer sparse AE. Their method was tested on data obtained from MC simulation in brain, Zubal phantoms, and also on real patient whole-body PET data. Their system outperformed the traditional MLEM algorithm and total variation regularization based method in noise reduction. The best results obtained were (signal to noise ratio (SNR), Bias, Variance) 70.42, 0.0365, 0.0071 and 68.41, 0.0369, 0.009 on the Brain and Zubal phantom respectively. It recovered more details and was independent of the data size. Related to the real patient data reconstruction results, their system produced a cleaner and smoother result, assessed by visual inspection.

Jiao and Ourselin (2017) proposed an alternative approach for PET reconstruction with multi-scale FCN (msfCNN). For each pixel from the back projection of sinograms, the model predicted a value in the reconstructed image or, in other words, the image reconstruction process was regarded as a dense prediction problem using msfCNN. The model was only trained using simulated data generated by brain scans in the ADNI database and tested on real oncological and neurodegenerative data.

To address the noise challenge characteristic of short dynamic scans, Spencer and Wang (2017) develop a new kernel method. The proposed dual kernel approach enhances traditional kernel-based dynamic PET image reconstruction techniques. The dual kernel combines a non-local kernel with a local convolutional kernel that can be trained from image priors. The approach is evaluated for shortened dynamic FDG-PET imaging using a digital brain phantom. Simulation results have demonstrated that the dual-kernel approach can achieve better image quality than both the standard reconstruction approach and the single kernel approach.

Kim et al. (2018) propose an iterative PET reconstruction using a CNN prior. The denoising CNN (DnCNN) method was used and the network was trained using regular dose images as ground-truth and low dose images reconstructed from downsampled data by Poisson thinning as input. A local linear fitting (LLF) function incorporated with DnCNN prior is used to improve the image quality by preventing bias. By using the LLF function, the method becomes robust to noise level changes. In the bias and variance studies, the method outperforms conventional methods.

Inspired by the deep image prior framework, Gong et al. (2018) propose a personalised network training method where no prior training pairs are needed, but only the patient's own prior information. The network is updated during the iterative reconstruction process using the patient specific prior information and measured data. The maximum likelihood estimation is formulated as a constrained optimisation problem and solved it using the alternating direction method of multipliers algorithm. MRI and PET reconstruction was employed as an example to demonstrate the effectiveness of the framework. Quantification results based on simulation and real data show that the proposed reconstruction framework can outperform Gaussian post-smoothing and anatomically-guided reconstructions using the kernel method or the neural network penalty. The same team, in Gong et al. (2019), also presents results for a lung patient study and a brain patient study.

Table 14 outlines the main characteristics (anatomic interest, dataset, and best performance results) of the previously discussed works on the PET modality.

**Conclusions** The presented studies used completely different approaches, datasets and evaluation metrics (Table 14). In general, the training process was not done with data of the same complexity as the application, using either simulated or very small datasets. The proposed models can be appropriately improved with more realistic training data. Furthermore, RNNs would be a pertinent DL technique to be exploited under this application since they are appropriate to use with sequential input data (in this case, a sequence of images).

#### 4.4.3 PET applications: estimating TOF

In this proof-of-concept study, Berg and Cherry (2018) evaluated the use of CNNs to estimate TOFs directly from coincident events. The waveform pairs were stored and labelled with the ground-truth TOFs, which were obtained through the distance between the positron emitting source and the coincident detectors. Two different configurations with a variable number of layers for CNN architecture were compared: the fixed configuration that used small convolutional filters and the tapered 6-layer configuration with the size of the filters progressively decreasing along the layers. They claimed to improve coincidence timing resolution up to 23% and demonstrated the greater impact of CNN depth (number of layers) instead of filter size and number of feature maps. The attraction of CNN to estimate TOF is that minimal pre-processing is required.

Table 14 outlines the main characteristics (anatomic interest, dataset, and best performance results) of the previously discussed work on PET modality.

**Conclusions** The analysed study validates the use of DL techniques, in particular, CNN, in a non-conventional application: the estimation of TOF. Although the input of this study is not directly PET images, the waveforms are derived from it and this study can have a direct impact on the PET modality with a promising improvement of coincidence timing resolution (Table 15).

#### 4.4.4 PET applications: generating PET

With the purpose of reducing the time and resources spent on extra exams, there is an interest in applying DL to convert images between different modalities (Ben-Cohen et al. 2017, 2019). Also, to cope with the well known need of DL techniques of a lot of data to train (Domingues et al. 2018), other works use data augmentation for the improvement of PET-based CAD systems (Bi et al. 2017a). Concerning the techniques, GAN is the most used architecture (Bi et al. 2017a; Ben-Cohen et al. 2017, 2019).

Bi et al. (2017a) developed a multi-channel GAN (M-GAN) to produce PET images that takes the annotations (labels) to synthesise high uptake regions and the corresponding CT images to constrain the appearance consistency. The Generator learns from the label and CT images to produce a synthetic PET, while the discriminator learns from synthetic and real PET. The dataset consisted of 50 lung cancer studies composed of a CT volume and a PET volume. Only trans-axial slices that contained tumours were considered. The multi-channel approach was compared to single channel approaches: using labels (LB-GAN) and using only CTs (CT-GAN) and achieved the best results, with the lowest MAE of 4.60 and the highest PSNR of 28.06. Furthermore, the M-GAN synthetic PET data created was used to train a FCN to detect tumors and achieved comparable results (precision = 81.73%, recall = 52.38%, f-score = 63.84%) to using real PET images (precision = 88.31%, recall = 55.17%, f-score = 66.38%), which shows that the data produced was realistic.

Ben-Cohen et al. have been publishing several works in the field (Ben-Cohen et al. 2017, 2019). Claiming to be the first work that explores CT to PET translation using DL, Ben-Cohen et al. (2017) presents a system where a blended image is produced from the results of a FCN and a GAN. An extension is then presented in Ben-Cohen et al. (2019) where instead of blending the resulting images, the output image of the FCN is refined by a cGAN. The dataset contains CT and PET pairs, constrained to the liver region. In Ben-Cohen et al. (2017), evaluation is based on the tumor detection performance, where 24 out of 26 tumours were detected at a cost of 2 FPs for the 8 test scans. In Ben-Cohen et al. (2019), reconstruction evaluation was made with MAE and PSNR, achieving average values of 0.72 and 30.22, respectively.

Table 16 outlines the main characteristics (anatomic interest, dataset, and best performance results) of the previously discussed work on PET integrated with CT modality.

**Conclusions** There is an interest in applying DL approaches to convert between different image modalities. The analysed studies aimed at converting CT images to PET images obtained encouraging results for the complexity of the engaged task (Table 16) that need more exploration to get a more secure system. However, the purpose of these studies can be regarded as ambitious since different modalities have different expressed lesions. Some lesions are only visible in one modality, so the translation of one modality to another is not straightforward.

#### 4.4.5 PET applications: radiation adaptation

It is known that the radiation can cause adverse effects and it is thus of interest to apply the least possible dose to the patients. The works reviewed in this section have as goal the development of tools to aid in the definition of the radiation to apply (Tseng et al. 2017) or to reconstruct high-quality images from low quality ones, where less radiation has been applied to the patient (Xiang et al. 2017).

Tseng et al. (2017) explored a multicomponent deep reinforcement learning to develop a dynamic clinical decision making in adaptive radiotherapy on historical treatment plans for NSCLC patients. Reinforcement learning aimed to mimic how clinicians decide on the dose fraction needed to be prescribed and are formulated as a Markov decision process. The dataset contained NSCLC of 114 patients who received radiotherapy. The main framework was composed of three neural networks components. First, a GANs was employed in order to generate more radiotherapy patients data due to the small available dataset. Second, a radiotherapy artificial environment was reconstructed by a DNN. Third, a deep Q-network was applied to search for the optimal policy sequence for response-based dose adaptation. They concluded that the deep reinforcement learning achieved similar results to real clinical decisions that were applied in an adaptive dose escalation clinical protocol and could be used to make recommendations to adjust the current treatment. Also, this approach needs extra validations on larger datasets.

Xiang et al. (2017) estimate high-quality standard dose PET (SPET) from the combination of low-quality low-dose PET (LPET) and the accompanying T1-weighted acquisition from MRI. CNNs are adapted to account for the two channel inputs of LPET and T1, and directly learn the end-to-end mapping between the inputs and the SPET output. Then, multiple CNN modules are integrated following the auto-context strategy, such that the tentatively estimated SPET of an early CNN can be iteratively refined by subsequent CNNs. The dataset contains 16 subjects, and performance is evaluated with NMSE and PSNR. The average PSNR scores are 24.76, while average NMSE scores are 0.0206.

Wang et al. (2018b) also estimates SPET. In order to achieve this, the authors use a 3D cGAN, where the model is conditioned on an input LPET and generates a corresponding output SPET. Specifically, to render the same underlying information between the LPET and SPET, a 3D U-net-like deep architecture that combines hierarchical features by using skip connection is designed as the generator network to synthesise the SPET. In order to guarantee the synthesised PET to be close to the real one, the estimation error loss in addition to the discriminator feedback is taken into consideration when training the generator network. Furthermore, a concatenated 3D cGAN based progressive refinement scheme is also proposed to further improve the quality of estimated images. Validation was done on a dataset with 8 normal subjects and 8 subjects diagnosed as MCI. The authors conclude that the adversarial training brings higher PSNR and lower NMSE for both the normal and the MCI subjects.

Table 17 outlines the main characteristics of the previously discussed works on the PET modality for radiation adaptation.

**Conclusions** To reduce the patient's exposure to radiation, while maintaining the high quality of the images is, without a doubt, a very important application. We foresee that the amount of works on this topic will substantially increase in a near future.

## 5 Discussion

Medical imaging analysis has benefited from the research in DL techniques. In our study, a comprehensive review of DL techniques applied to two different medical exam modalities was included, as well as their combination. Concerning CT, DL techniques were used for segmentation, classification, detection, pathology stage and treatment assessment, and noise and artefacts reduction. The relative availability of CT data has likely promoted DL progress in this modality. With regard to PET images, DL techniques were used for classification, prediction of response to treatment, reconstruction, and estimating TOF. There are a limited amount of work directed to PET. The majority of the revised works used private datasets, which prevent the replication and further improvement of results by other researchers. Also, the number of patients enrolled in most of these studies can be considered small.

In terms of dimensionality, we can categorise DL schemes that 2D or 3D data as input. Most of the (pre-trained) CNNs are designed for 2D natural images. They might be not well suited for medical imaging analysis, since some medical images are 3D volumetric images, such as CT and PET. Compared to 2D CNN, 3D CNN with 3D convolution operations, can better model and preserve the spatial neighbourhood and the 3D spatial information.

DL techniques have been proposed as an alternative to traditional feature-based approaches. But the change conducted by DL has not made feature-based approaches discarded. There are cases where these traditional approaches or the combination of both approaches obtained better results.

Despite showing encouraging results, there are multiple drawbacks to be resolved before the introduction of DL techniques into clinical workflow. These techniques are highly influenced by the quality and amount of training data. They require a large number of training images for determining a large number of parameters which directly implies a high computational cost. Not only the unavailability of image data but also relevant labelling for these images are a current problem. The tendency of overfitting is also a factual issue. Considering the differences on imaging acquisitions, the evaluation of methods requires progress, in order to prove their prevalent usefulness. Often classification is presented as a binary task which is a simplification of the reality in medical imaging analysis. A solution is to turn the DL approach into a multi-class system. Still related to data, the class imbalance is a common problem. The detection of rare diseases is a typical example, where very few samples exist to represent the disease cases. A system developed with an imbalance dataset may have the consequence of assigning all items to a single majority class. It apparently outperformed other approaches related to evaluation measures that do not take into account imbalanced classes. Furthermore, interpretability is an important concept in the healthcare context. As such, the black box nature of the current DL technique is a barrier. If the clinician cannot validate the approach and understand the logical bases of the method, it will certainly not be accepted.

Biomedical researchers may benefit from this survey for inspiration in future research for CT and PET. It is expected that DL with image input will be the mainstream technology in medical imaging in the next few years. As such, it is predictable that the design of novel DL techniques will happen.

**Acknowledgements** This article is a result of the project NORTE-01-0145-FEDER-000027, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

## References

- Afshari S, BenTaieb A, Hamarneh G (2018) Automatic localization of normal active organs in 3D PET scans. *Comput Med Imaging Graph* 70:111–118. <https://doi.org/10.1016/j.compmedimag.2018.09.008>
- Ait Skourt B, El Hassani A, Majda A (2018) Lung CT image segmentation using deep neural networks. *Procedia Comput Sci* 127:109–113. <https://doi.org/10.1016/j.procs.2018.01.104>
- Ali M, Son DH, Kang SH, Nam SR (2017) An accurate CT saturation classification using a deep learning approach based on unsupervised feature extraction and supervised fine-tuning strategy. *Energies* 10(11):1830. <https://doi.org/10.3390/en10111830>
- Antonio CB, Bautista LGC, Labao AB, Naval PC (2018) Vertebra fracture classification from 3D CT lumbar spine segmentation masks using a convolutional neural network. In: Asian conference on intelligent information and database systems. Lecture notes in artificial intelligence, vol 10752. Springer, pp 449–458. [https://doi.org/10.1007/978-3-319-75420-8\\_43](https://doi.org/10.1007/978-3-319-75420-8_43)
- Asadi S, Abdullah R, Yah Y, Nazir S (2019) Understanding institutional repository in higher learning institutions: a systematic literature review and directions for future research. *IEEE Access* 7:35242–35263. <https://doi.org/10.1109/ACCESS.2019.2897729>
- Barbu A, Lu L, Roth H, Seff A, Summers RM (2018) An analysis of robust cost functions for CNN in computer-aided diagnosis. *Comput Methods Biomech Biomed Eng Imaging Vis* 6(3):253–258. <https://doi.org/10.1080/21681163.2016.1138240>
- Belharbi S, Chatelain C, Héralut R, Adam S, Thureau S, Chastan M, Modzelewski R (2017) Spotting L3 slice in CT scans using deep convolutional network and transfer learning. *Comput Biol Med* 87(May):95–103. <https://doi.org/10.1016/j.compbiomed.2017.05.018>
- Ben-Cohen A, Klang E, Raskin SP, Amitai MM, Greenspan H (2017) Virtual PET images from CT data using deep convolutional networks: initial results. In: Lecture notes in computer science, vol 10557. Springer, pp 49–57. [https://doi.org/10.1007/978-3-319-68127-6\\_6](https://doi.org/10.1007/978-3-319-68127-6_6)
- Ben-Cohen A, Klang E, Raskin SP, Soffer S, Ben-Haim S, Konen E, Amitai MM, Greenspan H (2019) Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. *Eng Appl Artif Intell* 78:186–194. <https://doi.org/10.1016/j.engappai.2018.11.013>
- Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35(8):1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>
- Berg E, Cherry SR (2018) Using convolutional neural networks to estimate time-of-flight from PET detector waveforms. *Phys Med Biol* 63(2):02LT01. <https://doi.org/10.1088/1361-6560/aa9dc5>
- Bessa S, Domingues I, Cardoso JS, Passarinho P, Cardoso P, Rodrigues V, Lage F (2014) Normal breast identification in screening mammography: a study on 18000 images. In: IEEE international conference on bioinformatics and biomedicine (BIBM), pp 325–330
- Bhatt J, Joshi M, Sharma M (2018) Early detection of lung cancer from CT images: nodule segmentation and classification using deep learning. In: Tenth international conference on machine vision (ICMV). SPIE, p 29. <https://doi.org/10.1117/12.2309530>
- Bi L, Kim J, Kumar A, Feng D, Fulham M (2017a) Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In: Lecture notes in computer science, vol 10555. Springer, Cham, pp 43–51. [https://doi.org/10.1007/978-3-319-67564-0\\_5](https://doi.org/10.1007/978-3-319-67564-0_5)
- Bi L, Kim J, Kumar A, Wen L, Feng D, Fulham M (2017b) Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies. *Comput Med Imaging Graph* 60:3–10. <https://doi.org/10.1016/j.compmedimag.2016.11.008>
- Bibault JE, Giraud P, Housset M, Durdux C, Taieb J, Berger A, Coriat R, Chaussade S, Dousset B, Nordlinger B, Burgun A (2018) Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep* 8(1):12611. <https://doi.org/10.1038/s41598-018-30657-6>
- Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO (2018) Automatic lesion detection and segmentation of 18F-FET PET in gliomas: a full 3D U-Net convolutional neural network study. *PLoS ONE* 13(4):e0195798. <https://doi.org/10.1371/journal.pone.0195798>
- Blau N, Klang E, Kiryati N, Amitai M, Portnoy O, Mayer A (2018) Fully automatic detection of renal cysts in abdominal CT scans. *Int J Comput Assist Radiol Surg* 13(7):957–966. <https://doi.org/10.1007/s11548-018-1726-6>
- Boas FE, Fleischmann D (2012) CT artifacts: causes and reduction techniques. *Imaging Med* 4(2):229–240. <https://doi.org/10.2217/iim.12.13>
- Brenner DJ, Hricak H (2010) Radiation exposure from medical imaging. *JAMA* 304(2):208. <https://doi.org/10.1001/jama.2010.973>
- Bushberg JT, Boone JM (2011) The essential physics of medical imaging. Lippincott Williams & Wilkins, Philadelphia

- Cha KH, Hadjiiski L, Samala RK, Chan HP, Caoili EM, Cohan RH (2016) Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Med Phys* 43(4):1882–1896. <https://doi.org/10.1118/1.4944498>
- Cha KH, Hadjiiski L, Chan HP, Weizer AZ, Alva A, Cohan RH, Caoili EM, Paramagul C, Samala RK (2017a) Bladder cancer treatment response assessment in CT using radiomics with deep-learning. *Sci Rep* 7(1):8738. <https://doi.org/10.1038/s41598-017-09315-w>
- Cha KH, Hadjiiski LM, Chan HP, Caoili EM, Cohan RH, Weizer A, Samala RK (2017b) Computer-aided detection of bladder masses in CT urography (CTU). In: *Medical imaging: computer-aided diagnosis, SPIE*, vol 10134, p 1013403. <https://doi.org/10.1117/12.2255668>
- Chen CM, Chou YH, Tagawa N, Do Y (2013) Computer-aided detection and diagnosis in medical imaging. *Comput Math Methods Med* 2013:1–2. <https://doi.org/10.1155/2013/790608>
- Chen YJ, Hua KL, Hsu CH, Cheng WH, Hidayati SC (2015) Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets Therapy*. <https://doi.org/10.2147/OTT.S80733>
- Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G (2017a) Low-dose CT with a residual encoder–decoder convolutional neural network. *IEEE Trans Med Imaging* 36(12):2524–2535. <https://doi.org/10.1109/TMI.2017.2715284>
- Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G (2017b) Low-dose CT via convolutional neural network. *Biomed Opt Express* 8(2):679. <https://doi.org/10.1364/BOE.8.000679>
- Chen Y, Ren Y, Fu L, Xiong J, Larsson R, Xu X, Sun J, Zhao J (2018) A 3D convolutional neural network framework for polyp candidates detection on the limited dataset of CT colonography. In: *40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp 678–681. <https://doi.org/10.1109/EMBC.2018.8512305>
- Cheng D, Liu M (2017) Combining convolutional and recurrent neural networks for Alzheimer’s disease diagnosis using PET images. In: *IEEE international conference on imaging systems and techniques (IST)*, pp 1–5. <https://doi.org/10.1109/IST.2017.8261461>
- Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, Huang CS, Shen D, Chen CM (2016) Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep* 6(1):24454. <https://doi.org/10.1038/srep24454>
- Chmelik J, Jakubicek R, Walek P, Jan J, Ourednicek P, Lambert L, Amadori E, Gavelli G (2018) Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Med Image Anal* 49:76–88. <https://doi.org/10.1016/j.media.2018.07.008>
- Chmelik J, Jakubicek R, Jan J, Ourednicek P, Lambert L, Amadori E, Gavelli G (2019) Fully automatic CAD system for segmentation and classification of spinal metastatic lesions in CT data. In: *World congress on medical physics and biomedical engineering 2018*, vol 68. Springer, pp 155–158. [https://doi.org/10.1007/978-981-10-9035-6\\_28](https://doi.org/10.1007/978-981-10-9035-6_28)
- Choi H, Jin KH (2018) Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav Brain Res* 344:103–109. <https://doi.org/10.1016/j.bbr.2018.02.017>
- Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Armbruster M, Hofmann F, D’Anastasi M, Sommer WH, Ahmadi SA, Menze BH (2016) Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3D conditional random fields. In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Springer, pp 415–423
- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Springer, pp 424–432. [https://doi.org/10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)
- Ciampi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, de Jong PA, Prokop M, van Ginneken B (2015) Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Med Image Anal* 26(1):195–202. <https://doi.org/10.1016/j.media.2015.08.001>
- Costello JE, Cecava ND, Tucker JE, Bau JL (2013) CT radiation dose: current controversies and dose reduction strategies. *Am J Roentgenol* 201(6):1283–1290. <https://doi.org/10.2214/AJR.12.9720>
- Cui J, Liu X, Wang Y, Liu H (2017) Deep reconstruction model for dynamic PET images. *PLoS ONE* 12(9):e0184667. <https://doi.org/10.1371/journal.pone.0184667>
- Czakov J, Drapejkowski F, Zurek G, Giedziun P, Zebrowski J, Dyrka W (2016) Machine learning methods for accurate delineation of tumors in PET images. *CoRR arXiv:1610.09493*
- Da C, Zhang H, Sang Y (2015) Brain CT image classification with deep neural networks. In: *18th Asia pacific symposium on intelligent and evolutionary systems*, vol 1, pp 653–662. [https://doi.org/10.1007/978-3-319-13359-1\\_50](https://doi.org/10.1007/978-3-319-13359-1_50)

- Das A, Acharya UR, Panda SS, Sabut S (2019) Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. *Cognit Syst Res* 54:165–175. <https://doi.org/10.1016/j.cogsys.2018.12.009>
- de Carvalho Filho AO, Silva AC, de Paiva AC, Nunes RA, Gattass M (2018) Classification of patterns of benignity and malignancy based on CT using topology-based phylogenetic diversity index and convolutional neural network. *Pattern Recognit* 81:200–212. <https://doi.org/10.1016/j.patcog.2018.03.032>
- de Vos BD, Wolterink JM, de Jong PA, Leiner T, Viergever MA, Išgum I (2017) ConvNet-based localization of anatomical structures in 3-D medical images. *IEEE Trans Med Imaging* 36(7):1470–1481. <https://doi.org/10.1109/TMI.2017.2673121>
- Deng CL, Jiang HY, Li HM (2017) Automated high uptake regions recognition and lymphoma detection based on fully convolutional networks on chest and abdomen PET image. In: International conference on medicine sciences and bioengineering (ICMSB), pp 331–336. <https://doi.org/10.12783/dtbb/icmsb2017/17985>
- Diniz JOB, Diniz PHB, Valente TLA, Silva AC, Paiva AC (2019) Spinal cord detection in planning CT for radiotherapy through adaptive template matching, IMSLIC and convolutional neural networks. *Comput Methods Programs Biomed* 170:53–67. <https://doi.org/10.1016/j.cmpb.2019.01.005>
- Doersch C (2016) Tutorial on variational autoencoders. *CoRR arXiv:1606.05908*
- Domingues III, Cardoso JS (2014) Max-ordinal learning. *IEEE Trans Neural Netw Learn Syst* 25(7):1384–1389. <https://doi.org/10.1109/TNNLS.2013.2287381>
- Domingues I, Abreu PH, Santos J (2018) Bi-rads classification of breast cancer: a new pre-processing pipeline for deep models training. In: 25th IEEE international conference on image processing (ICIP), pp 1378–1382. <https://doi.org/10.1109/ICIP.2018.8451510>
- Domingues I, Sampaio I, Duarte H, dos Santos JAM, Abreu PH (2019) Computer vision in esophageal cancer: a literature review. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2019.2930891>
- Domingues I, Cardoso JS (2013) Mass detection on mammogram images: a first assessment of deep learning techniques. In: 19th Portuguese conference on pattern recognition
- Dormer JD, Halicek M, Ma L, Reilly CM, Fei B, Schreibmann E (2018) Convolutional neural networks for the detection of diseased hearts using CT images and left atrium patches. In: Medical imaging: computer-aided diagnosis, SPIE, p 107. <https://doi.org/10.1117/12.2293548>
- Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng PA (2016) 3D deeply supervised network for automatic liver segmentation from CT volumes. In: International conference on medical image computing and computer-assisted intervention (MICCAI). Springer, pp 149–157
- Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, Heng PA (2017) 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal* 41:40–54. <https://doi.org/10.1016/j.media.2017.05.001>
- Erhan D, Manzagol PA, Bengio Y, Bengio S, Vincent P (2009) The difficulty of training deep architectures and the effect of unsupervised pre-training. In: Proceedings of the international conference on artificial intelligence and statistics, vol 5, pp 153–160
- Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P, Bengio S (2010) Why does unsupervised pre-training help deep learning? *J Mach Learn Res* 11:625–660
- Erickson ZA, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ (2017) Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging* 30(4):449–459. <https://doi.org/10.1007/s10278-017-9983-4>
- Farag A, Lu L, Roth HR, Liu J, Turkbey E, Summers RM (2017) A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. *IEEE Trans Image Process* 26(1):386–399. <https://doi.org/10.1109/TIP.2016.2624198>
- Fechter T, Adebahr S, Baltas D, Ben Ayed I, Desrosiers C, Dolz J (2017) Esophagus segmentation in CT via 3D fully convolutional neural network and random walk. *Med Phys* 44(12):6341–6352. <https://doi.org/10.1002/mp.12593>
- Feng A, Chen Z, Wu X, Ma Z (2017) From convolutional to recurrent: case study in Nasopharyngeal Carcinoma segmentation. In: International conference on the frontiers and advances in data science (FADS). IEEE, pp 18–22. <https://doi.org/10.1109/FADS.2017.8253187>
- Fischer A, Igel C (2014) Training restricted Boltzmann machines: an introduction. *Pattern Recognit* 47(1):25–39. <https://doi.org/10.1016/j.patcog.2013.05.025>
- Fukushima K, Miyake S (1982) Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. Springer, pp 267–285. [https://doi.org/10.1007/978-3-642-46466-9\\_18](https://doi.org/10.1007/978-3-642-46466-9_18)
- Gao XW, Hui R (2016) A deep learning based approach to classification of CT brain images. In: IEEE SAI computing conference (SAI), pp 28–31. <https://doi.org/10.1109/SAI.2016.7555958>

- Gao M, Xu Z, Lu L, Wu A, Noguez I, Summers RM, Mollura DJ (2016) Segmentation label propagation using deep convolutional neural networks and dense conditional random field. In: IEEE 13th international symposium on biomedical imaging (ISBI), pp 1265–1268. <https://doi.org/10.1109/ISBI.2016.7493497>
- Gao XW, Hui R, Tian Z (2017) Classification of CT brain images based on deep learning networks. *Comput Methods Programs Biomed* 138:49–56. <https://doi.org/10.1016/j.cmpb.2016.10.007>
- Gao M, Bagci U, Lu L, Wu A, Buty M, Shin HC, Roth H, Papadakis GZ, Depeursing A, Summers RM, Xu Z, Mollura DJ (2018) Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis* 6(1):1–6. <https://doi.org/10.1080/21681163.2015.1124249>
- Gerard SE, Patton TJ, Christensen GE, Bayouth JE, Reinhardt JM (2019) FissureNet: a deep learning approach for pulmonary fissure detection in CT images. *IEEE Trans Med Imaging* 38(1):156–166. <https://doi.org/10.1109/TMI.2018.2858202>
- Ghesu FC, Georgescu B, Zheng Y, Grbic S, Maier A, Hornegger J, Comaniciu D (2019) Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. *IEEE Trans Pattern Anal Mach Intell* 41(1):176–189. <https://doi.org/10.1109/TPAMI.2017.2782687>
- Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira SP, Clarkson MJ, Barratt DC (2018) Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Trans Med Imaging* 37(8):1822–1834. <https://doi.org/10.1109/TMI.2018.2806309>
- Golan R, Jacob C, Denzinger J (2016) Lung nodule detection in CT images using deep convolutional neural networks. In: IEEE international joint conference on neural networks (IJCNN), pp 243–250. <https://doi.org/10.1109/IJCNN.2016.7727205>
- Gong K, Catana C, Qi J, Li Q (2018) PET image reconstruction using deep image prior. *IEEE Trans Med Imaging*. <https://doi.org/10.1109/TMI.2018.2888491>
- Gong K, Guan J, Kim K, Zhang X, Yang J, Seo Y, El Fakhri G, Qi J, Li Q (2019) Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans Med Imaging* 38(3):675–685. <https://doi.org/10.1109/TMI.2018.2869871>
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems 27. Curran Associates Inc., pp 2672–2680
- Gordon M, Hadjiiski L, Cha K, Chan HP, Samala R, Cohan RH, Caoili EM (2017) Segmentation of inner and outer bladder wall using deep-learning convolutional neural network in CT urography. In: Medical imaging: computer-aided diagnosis, SPIE, vol 10134. <https://doi.org/10.1117/12.2255528>
- Greenspan H, van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 35(5):1153–1159. <https://doi.org/10.1109/TMI.2016.2553401>
- Grewal M, Srivastava MM, Kumar P, Varadarajan S (2018) RADnet: radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In: IEEE 15th international symposium on biomedical imaging (ISBI), pp 281–284. <https://doi.org/10.1109/ISBI.2018.8363574>
- Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T (2018) Recent advances in convolutional neural networks. *Pattern Recognit* 77:354–377. <https://doi.org/10.1016/j.patcog.2017.10.013>
- Guo Z, Li X, Huang H, Guo N, Li Q (2018) Medical image segmentation based on multi-modal convolutional neural network: study on image fusion schemes. In: IEEE 15th international symposium on biomedical imaging (ISBI), pp 903–907. <https://doi.org/10.1109/ISBI.2018.8363717>
- Hamidian S, Sahiner B, Petrick N, Pezeshk A (2017) 3D convolutional neural network for automatic detection of lung nodules in chest CT. In: Medical imaging: computer-aided diagnosis, SPIE, vol 10134, p 1013409. <https://doi.org/10.1117/12.2255795>
- Hashimoto N, Suzuki K, Liu J, Hirano Y, MacMahon H, Kido S (2018) Deep neural network convolution (NNC) for three-class classification of diffuse lung disease opacities in high-resolution CT (HRCT): consolidation, ground-glass opacity (GGO), and normal opacity. In: Medical imaging: computer-aided diagnosis, SPIE, p 113. <https://doi.org/10.1117/12.2293550>
- Hastie T, Friedman J, Tibshirani R (2001) The elements of statistical learning, 2nd edn. Springer, New York. <https://doi.org/10.1007/978-0-387-21606-5>
- Hatt M, Laurent B, Ouahabi A, Fayad H, Tan S, Li L, Lu W, Jaouen V, Tauber C, Czakon J, Drapejkowski F, Dyrka W, Camarasu-Pop S, Cervenansky F, Girard P, Glatar T, Kain M, Yao Y, Barillot C, Kirov A, Visvikis D (2018) The first MICCAI challenge on PET tumor segmentation. *Med Image Anal* 44:177–195. <https://doi.org/10.1016/j.media.2017.12.007>
- He J, Yang Y, Wang Y, Zeng D, Bian Z, Zhang H, Sun J, Xu Z, Ma J (2019) Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction. *IEEE Trans Med Imaging* 38(2):371–382. <https://doi.org/10.1109/TMI.2018.2865202>

- Hong HA, Sheikh UU (2016) Automatic detection, segmentation and classification of abdominal aortic aneurysm using deep learning. In: IEEE 12th international colloquium on signal processing and its applications (CSPA), pp 242–246. <https://doi.org/10.1109/CSPA.2016.7515839>
- Huang L, Xia W, Zhang B, Qiu B, Gao X (2017) MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images. *Comput Methods Programs Biomed* 143:67–74. <https://doi.org/10.1016/j.cmpb.2017.02.013>
- Huang B, Chen Z, Wu PM, Ye Y, Feng ST, Wong CYO, Zheng L, Liu Y, Wang T, Li Q, Huang B (2018) Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: a dual-center study. *Contrast Media Mol Imaging*. <https://doi.org/10.1155/2018/8923028>
- Humpire Mamani GE, Setio AAA, van Ginneken B, Jacobs C (2017) Organ detection in thorax abdomen CT using multi-label convolutional neural networks. In: *Medical imaging: computer-aided diagnosis, SPIE*, vol 10134, p 1013416. <https://doi.org/10.1117/12.2254349>
- Hu Z, Muhammad A, Zhu M (2018) Pulmonary nodule detection in CT images via deep neural network. In: 2nd International conference on graphics and signal processing (ICGSP). ACM Press, pp 79–83. <https://doi.org/10.1145/3282286.3282302>
- Hussein S, Cao K, Song Q, Bagci U (2017) Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: International conference on information processing in medical imaging, pp 249–260. [https://doi.org/10.1007/978-3-319-59050-9\\_20](https://doi.org/10.1007/978-3-319-59050-9_20)
- Ibragimov B, Xing L (2017) Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys* 44(2):547–557. <https://doi.org/10.1002/mp.12045>
- Ibragimov B, Toesca D, Chang D, Koong A, Xing L (2017) Combining deep learning with anatomy analysis for segmentation of portal vein for liver SBRT planning. *Phys Med Biol* 62(23):8943–8958. <https://doi.org/10.1088/1361-6560/aa9262>
- Işgum I, van Ginneken B, Lessmann N (2018) Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images. In: *Medical imaging: image processing, SPIE*, vol 10574, p 7. <https://doi.org/10.1117/12.2292731>
- Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 5967–5976. <https://doi.org/10.1109/CVPR.2017.632>
- Jackson P, Hardcastle N, Dawe N, Kron T, Hofman MS, Hicks RJ (2018) Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy. *Front Oncol*. <https://doi.org/10.3389/fonc.2018.00215>
- Jiao J, Ourselin S (2017) Fast PET reconstruction using multi-scale fully convolutional neural networks. *CoRR* [arXiv:1704.07244](https://arxiv.org/abs/1704.07244)
- Jin X, Ye H, Li L, Xia Q (2017) Image segmentation of liver CT based on fully convolutional network. In: IEEE 10th international symposium on computational intelligence and design (ISCID), pp 210–213. <https://doi.org/10.1109/ISCID.2017.49>
- Jnawali K, Arbabshirani MR, Rao N, Patel AA (2018) Deep 3D convolution neural network for CT brain hemorrhage classification. In: *Medical imaging: computer-aided diagnosis, SPIE*, p 47. <https://doi.org/10.1117/12.2293725>
- Jung Kh, Park H, Hwang W (2017) Deep learning for medical image analysis: applications to computed tomography and magnetic resonance imaging. *Hanyang Med Rev* 37(2):61. <https://doi.org/10.7599/hmr.2017.37.2.61>
- Jung H, Kim B, Lee I, Lee J, Kang J (2018) Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method. *BMC Med Imaging* 18(1):48. <https://doi.org/10.1186/s12880-018-0286-0>
- Kalinovsky A, Liauchuk V, Tarasau A (2017) Lesion detection in CT images using deep learning semantic segmentation technique. In: International archives of the photogrammetry, remote sensing and spatial information sciences (ISPRS), vol XLII-2/W4, pp 13–17. <https://doi.org/10.5194/isprs-archives-XLII-2-W4-13-2017>
- Kang E, Min J, Ye JC (2017) A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys* 44(10):e360–e375. <https://doi.org/10.1002/mp.12344>
- Kasban H, El-Bendary MAM, Salama DH (2015) A comparative study of medical imaging techniques. *Int J Inf Sci Intell Syst (IJISIS)* 4:37–58
- Keshavamurthy KN, Leary OP, Merck LH, Kimia B, Collins S, Wright DW, Allen JW, Brock JF, Merck D (2017) Machine learning algorithm for automatic detection of CT-identifiable hyperdense lesions associated with traumatic brain injury. In: *Medical imaging: computer-aided diagnosis, SPIE*, 1, p 10134-2G. <https://doi.org/10.1117/12.2254227>

- Kim BC, Sung YS, Suk HI (2016) Deep feature learning for pulmonary nodule classification in a lung CT. In: IEEE 4th international winter conference on brain-computer interface (BCI), pp 1–3. <https://doi.org/10.1109/IWW-BCI.2016.7457462>
- Kim K, Wu D, Gong K, Dutta J, Kim JH, Son YD, Kim HK, El Fakhri G, Li Q (2018) Penalized PET reconstruction using deep learning prior and local linear fitting. *IEEE Trans Med Imaging* 37(6):1478–1487. <https://doi.org/10.1109/TMI.2018.2832613>
- Kingma DP, Welling M (2014) Stochastic gradient VB and the variational auto-encoder. In: 2nd International conference on learning representations (ICLR)
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: 25th International conference on neural information processing systems (NIPS). Curran Associates Inc., USA, pp 1097–1105
- Kumar A, Fulham MJ, Feng D, Kim J (2018) Co-learning feature fusion maps from PET-CT images of lung cancer. *CoRR* [arXiv:1810.02492](https://arxiv.org/abs/1810.02492)
- Kumar D, Wong A, Clausi DA (2015) Lung nodule classification using deep features in CT images. In: IEEE 12th conference on computer and robot vision, pp 133–138. <https://doi.org/10.1109/CRV.2015.25>
- Lakshmanaprabu S, Mohanty SN, Shankar K, Arunkumar N, Ramirez G (2019) Optimal deep learning model for classification of lung cancer on CT images. *Future Gener Comput Syst* 92:374–382. <https://doi.org/10.1016/j.future.2018.10.009>
- Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, Kim N (2017) Deep learning in medical imaging: general overview. *Korean J Radiol* 18(4):570. <https://doi.org/10.3348/kjr.2017.18.4.570>
- Lee H, Hong H, Kim J, Jung DC (2018) Deep feature classification of angiomyolipoma without visible fat and renal cell carcinoma in abdominal contrast-enhanced CT images with texture image patches and hand-crafted feature concatenation. *Med Phys* 45(4):1550–1561. <https://doi.org/10.1002/mp.12828>
- Lee D, Choi S, Kim HJ (2019) High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains. *Med Phys* 46(1):104–115. <https://doi.org/10.1002/mp.13258>
- Lei L, Zhu H, Gong Y, Cheng Q (2018) A deep residual networks classification algorithm of fetal heart CT images. In: IEEE international conference on imaging systems and techniques (IST), pp 1–4. <https://doi.org/10.1109/IST.2018.8577179>
- Li Q, Nishikawa RM (2015) Computer-aided detection and diagnosis in medical imaging. Taylor & Francis, New York
- Lian C, Li H, Vera P, Ruan S (2018) Unsupervised co-segmentation of tumor in PET-CT images using belief functions based fusion. In: IEEE 15th international symposium on biomedical imaging (ISBI), pp 220–223. <https://doi.org/10.1109/ISBI.2018.8363559>
- Lian C, Ruan S, Denoeux T, Li H, Vera P (2019) Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions. *IEEE Trans Image Process* 28(2):755–766. <https://doi.org/10.1109/TIP.2018.2872908>
- Lin EC (2010) Radiation risk from medical imaging. *Mayo Clin Proc* 85(12):1142–1146. <https://doi.org/10.4065/mcp.2010.0260>
- Lin J, Kligerman S, Goel R, Sajedi P, Suntharalingam M, Chuong MD (2015) State-of-the-art molecular imaging in esophageal cancer management: implications for diagnosis, prognosis, and treatment. *J Gastrointest Oncol* 6(1):3. <https://doi.org/10.3978/j.issn.2078-6891.2014.062>
- Lisowska A, Beveridge E, Muir K, Poole I (2017) Thrombus detection in CT brain scans using a convolutional neural network. In: 10th International joint conference on biomedical engineering systems and technologies. SciTePress: Science and Technology Publications, pp 24–33. <https://doi.org/10.5220/0006114600240033>
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42(1995):60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Liu S (2018) Automated analysis of quantitative image biomarkers from low-dose chest CT scans. Doctor of philosophy, Cornell University. <https://doi.org/10.7298/X4VM49JN>
- Liu J, Wang S, Linguraru MG, Yao J, Summers RM (2015) Computer-aided detection of exophytic renal lesions on non-contrast CT images. *Med Image Anal* 19(1):15–29. <https://doi.org/10.1016/j.media.2014.07.005>
- Liu J, Lay N, Wei Z, Lu L, Kim L, Turkbey E, Summers RM (2016) Colitis detection on abdominal CT scans by rich feature hierarchies. In: *Medical imaging: computer-aided diagnosis, SPIE*, vol 9785, p 97851N. <https://doi.org/10.1117/12.2217681>
- Liu J, Lu L, Yao J, Bagheri M, Summers RM (2017) Pelvic artery calcification detection on CT scans using convolutional neural networks. In: *Medical imaging: computer-aided diagnosis, SPIE*, vol 10134, p 101341A. <https://doi.org/10.1117/12.2255247>

- Liu M, Cheng D, Yan W (2018a) Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front Neuroinform.* <https://doi.org/10.3389/fninf.2018.00035>
- Liu X, Hou F, Qin H, Hao A (2018b) Multi-view multi-scale CNNs for lung nodule type classification from CT images. *Pattern Recognit* 77:262–275. <https://doi.org/10.1016/j.patcog.2017.12.022>
- Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, van Elmpst W, Dekker A (2018) Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol* 126(2):312–317. <https://doi.org/10.1016/j.radonc.2017.11.012>
- Lyu J, Ling SH (2018) Using multi-level convolutional neural network for classification of lung nodules on CT images. In: 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 686–689. <https://doi.org/10.1109/EMBC.2018.8512376>
- Ma L, Guo R, Zhang G, Tade F, Schuster DM, Nieh P, Master V, Fei B (2017) Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion. In: Medical imaging: image processing, SPIE, vol 10133, p 101332O. <https://doi.org/10.1117/12.2255755>
- Maier J, Eulig E, Vöth T, Knaup M, Kuntz J, Sawall S, Kachelrieß M (2019) Real-time scatter estimation for medical CT using the deep scatter estimation: method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. *Med Phys* 46(1):238–249. <https://doi.org/10.1002/mp.13274>
- Martins P, Carbone I, Silva A, Teixeira A (2007) An MRI study of European Portuguese nasals. In: 8th Annual conference of the international speech communication association (Interspeech), C, pp 58–61
- Masood A, Sheng B, Li P, Hou X, Wei X, Qin J, Feng D (2018) Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *J Biomed Inform* 79:117–128. <https://doi.org/10.1016/j.jbi.2018.01.005>
- McCullough CH, Leng S, Yu L, Fletcher JG (2015) Dual- and multi-energy CT: principles, technical approaches, and clinical applications. *Radiology* 276(3):637–653. <https://doi.org/10.1148/radiol.2015142631>
- Men K, Dai J, Li Y (2017) Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys* 44(12):6377–6389. <https://doi.org/10.1002/mp.12602>
- Miki Y, Muramatsu C, Hayashi T, Zhou X, Hara T, Katsumata A, Fujita H (2017) Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med* 80:24–29. <https://doi.org/10.1016/j.compbimed.2016.11.003>
- Moses WW (2011) Fundamental limits of spatial resolution in PET. *Nucl Instrum Methods Phys Res Sect A Accel Spectrom Detect Assoc Equip* 648:S236–S240. <https://doi.org/10.1016/j.nima.2010.11.092>
- Näppi JJ, Pickhardt P, Kim DH, Hironaka T, Yoshida H (2017) Deep learning of contrast-coated serrated polyps for computer-aided detection in CT colonography. In: Medical imaging: computer-aided diagnosis, SPIE, vol 10134, p 101340H. <https://doi.org/10.1117/12.2255634>
- Nardelli P, Jimenez-Carretero D, Bermejo-Pelaez D, Ledesma-Carbayo MJ, Rahaghi FN, Estepar RSJ (2017) Deep-learning strategy for pulmonary artery-vein classification of non-contrast CT images. In: IEEE 14th international symposium on biomedical imaging (ISBI), pp 384–387. <https://doi.org/10.1109/ISBI.2017.7950543>
- Nardelli P, Jimenez-Carretero D, Bermejo-Pelaez D, Washko GR, Rahaghi FN, Ledesma-Carbayo MJ, San Jose Estepar R (2018) Pulmonary artery-vein classification in CT images using deep learning. *IEEE Trans Med Imaging* 37(11):2428–2440. <https://doi.org/10.1109/TMI.2018.2833385>
- Nazir S, Shahzad S, Mukhtar N (2019) Software birthmark design and estimation: a systematic literature. *Arab J Sci Eng.* <https://doi.org/10.1007/s13369-019-03718-9>
- Nogueira MA, Abreu PH, Martins P, Machado P, Duarte H, Santos J (2017) Image descriptors in radiology images: a systematic review. *Artif Intell Rev* 47(4):531–559. <https://doi.org/10.1007/s10462-016-9492-8>
- Oda H, Bhatia KK, Roth HR, Oda M, Kitasaka T, Iwano S, Homma H, Takabatake H, Mori M, Natori H, Schnabel JA, Mori K (2018) Dense volumetric detection and segmentation of mediastinal lymph nodes in chest CT images. In: Medical imaging: computer-aided diagnosis, SPIE, vol 10575, p 1. <https://doi.org/10.1117/12.2287066>
- Öman O, Mäkelä T, Salli E, Savolainen S, Kangasniemi M (2019) 3D convolutional neural networks applied to CT angiography in the detection of acute ischemic stroke. *Eur Radiol Exp* 3(1):8. <https://doi.org/10.1186/s41747-019-0085-6>

- Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, Imaizumi K, Fujita H (2019) Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. *BioMed Res Int* 2019:1–9. <https://doi.org/10.1155/2019/6051939>
- Özsavaş EE, Telatar Z, Dirican B, Sağır Ö, Beyzadeoğlu M (2014) Automatic segmentation of anatomical structures from CT scans of thorax for RTP. *Comput Math Methods Med*. <https://doi.org/10.1155/2014/472890>
- Patel A, van de Leemput SC, Prokop M, van Ginneken B, Manniesing R (2017) Automatic cerebrospinal fluid segmentation in non-contrast CT images using a 3D convolutional network. In: *Medical imaging: computer-aided diagnosis, SPIE*, vol 10134, p 1013420. <https://doi.org/10.1117/12.2254022>
- Paul R, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, Goldgof DB (2016) Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography* 2(4):388–395. <https://doi.org/10.18383/j.tom.2016.00211>
- Peng L, Lin L, Hu H, Li H, Ling X, Wang D, Han X, Iwamoto Y, Chen YW (2018) Classification of pulmonary emphysema in CT images based on multi-scale deep convolutional neural networks. In: 25th IEEE international conference on image processing (ICIP), pp 3119–3123. <https://doi.org/10.1109/ICIP.2018.8451514>
- Pereira G (2018) Deep learning techniques for the evaluation of response to treatment in Hodgkin Lymphoma. Master in biomedical engineering, University of Coimbra
- Pereira G, Domingues I, Martins P, Abreu PH, Duarte H, Santos J (2018) Registration of CT with PET: a comparison of intensity-based approaches. In: *International workshop on combinatorial image analysis (IWCIA)*
- Prassopoulos VK, Efthymiadou RD (2016) General principles of PET/CT imaging. In: *PET/CT in lymphomas*. Springer, Cham, pp 21–22. [https://doi.org/10.1007/978-3-319-27380-8\\_2](https://doi.org/10.1007/978-3-319-27380-8_2)
- Qiang Y, Ge L, Zhao X, Zhang X, Tang X (2017) Pulmonary nodule diagnosis using dual-modal supervised autoencoder based on extreme learning machine. *Expert Syst* 34(6):e12224. <https://doi.org/10.1111/exsy.12224>
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, pp 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Springer, pp 234–241
- Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, Liu J, Turkbey E, Summers RM (2014) A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Lecture notes in computer science, vol 8673. Springer, pp 520–527. [https://doi.org/10.1007/978-3-319-10404-1\\_65](https://doi.org/10.1007/978-3-319-10404-1_65)
- Roth HR, Farag A, Lu L, Turkbey EB, Summers RM (2015a) Deep convolutional networks for pancreas segmentation in CT imaging. In: *Medical imaging: image processing, SPIE*, vol 9413, p 94131G. <https://doi.org/10.1117/12.2081420>, [arXiv:1504.03967](https://arxiv.org/abs/1504.03967)
- Roth HR, Lee CT, Shin HC, Seff A, Kim L, Yao J, Lu L, Summers RM (2015b) Anatomy-specific classification of medical images using deep convolutional nets. In: *IEEE 12th international symposium on biomedical imaging (ISBI)*, pp 101–104. <https://doi.org/10.1109/ISBI.2015.7163826>
- Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, Summers RM (2015c) Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Springer, pp 556–564
- Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM (2016) Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans Med Imaging* 35(5):1170–1181. <https://doi.org/10.1109/TMI.2015.2482920>
- Sathyan H, Panicker JV (2018) Lung nodule classification using deep ConvNets on CT images. In: *IEEE 9th international conference on computing, communication and networking technologies (ICCCNT)*, pp 1–5. <https://doi.org/10.1109/ICCCNT.2018.8494084>
- Schwyzler M, Ferraro DA, Muehlethaler UJ, Curioni-Fontecedro A, Huellner MW, von Schulthess GK, Kaufmann PA, Burger IA, Messerli M (2018) Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks: initial results. *Lung Cancer* 126:170–173. <https://doi.org/10.1016/j.lungcan.2018.11.001>
- Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, Wille MMW, Naqibullah M, Sanchez CI, van Ginneken B (2016) Pulmonary nodule detection in CT images: false positive reduction using multi-view

- convolutional networks. *IEEE Trans Med Imaging* 35(5):1160–1169. <https://doi.org/10.1109/TMI.2016.2536809>
- Shaffie A, Soliman A, Ghazal M, Taher F, Dunlap N, Wang B, Van Berkel V, Gimelfarb G, Elmaghrawy A, El-Baz A (2018) A novel autoencoder-based diagnostic system for early assessment of lung cancer. In: 25th IEEE international conference on image processing (ICIP), pp 1393–1397. <https://doi.org/10.1109/ICIP.2018.8451595>
- Shaish H, Mutasa S, Makkari J, Chang P, Schwartz L, Ahmed F (2019) Prediction of lymph node maximum standardized uptake value in patients with cancer using a 3D convolutional neural network: a proof-of-concept study. *Am J Roentgenol* 212(2):238–244. <https://doi.org/10.2214/AJR.18.20094>
- Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, Cong W, Wang G (2018) 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. *IEEE Trans Med Imaging* 37(6):1522–1534. <https://doi.org/10.1109/TMI.2018.2832217>
- Sharma N, Ray A, Shukla K, Sharma S, Pradhan S, Srivastava A, Aggarwal L (2010) Automated medical image segmentation techniques. *J Med Phys* 35(1):3. <https://doi.org/10.4103/0971-6203.58777>
- Shen W, Yang F, Mu W, Yang C, Yang X, Tian J (2015a) Automatic localization of vertebrae based on convolutional neural networks. In: *Medical imaging: image processing, international society for optics and photonics*, vol 9413, p 94132E. <https://doi.org/10.1117/12.2081941>
- Shen W, Zhou M, Yang F, Yang C, Tian J (2015b) Multi-scale convolutional neural networks for lung nodule classification. In: *International conference on information processing in medical imaging*. Springer, pp 588–599. [https://doi.org/10.1007/978-3-319-19992-4\\_46](https://doi.org/10.1007/978-3-319-19992-4_46)
- Shen W, Zhou M, Yang F, Dong D, Yang C, Zang Y, Tian J (2016) Learning from experts: developing transferable deep features for patient-level lung cancer prediction. In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*. Springer, pp 124–131. [https://doi.org/10.1007/978-3-319-46723-8\\_15](https://doi.org/10.1007/978-3-319-46723-8_15)
- Shen D, Wu G, Hi Suk (2017a) Deep learning in medical image analysis. *Annu Rev Biomed Eng* 19(1):221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>
- Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y, Tian J (2017b) Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit* 61(SI):663–673. <https://doi.org/10.1016/j.patcog.2016.05.029>
- Shen S, Han SX, Aberle DR, Bui AA, Hsu W (2019) An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. *Expert Syst Appl* 128:84–95. <https://doi.org/10.1016/j.eswa.2019.01.048>
- Shi Y, Yang W, Gao Y, Shen D (2017) Does manual delineation only provide the side information in CT prostate segmentation? In: *International conference on medical image computing and computer-assisted intervention (MICCAI)*, vol 10435, pp 692–700. [https://doi.org/10.1007/978-3-319-66179-7\\_79](https://doi.org/10.1007/978-3-319-66179-7_79)
- Shin HC, Roth HR, Gao M, Lu L, Xu Z, Noguees I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35(5):1285–1298. <https://doi.org/10.1109/TMI.2016.2528162>
- Singh S, Srivastava A, Mi L, Chen K, Wang Y, Caselli RJ, Goradia D, Reiman EM (2017) Deep-learning-based classification of FDG-PET data for Alzheimer's disease categories. In: *13th International conference on medical information processing and analysis, international society for optics and photonics, SPIE*, vol 10572, p 84. <https://doi.org/10.1117/12.2294537>
- Smith RL, Florence V, Paisey S, Fittock E, Siebzehrubl F, Spezi E, Marshall C (2018) Deep learning pre-clinical medical image segmentation for automated organ-wise delineation of PET. *Eur J Nucl Med Mol Imaging* 45(1):S290
- Song Q, Zhao L, Luo X, Dou X (2017) Using deep learning for classification of lung nodules on computed tomography images. *J Healthc Eng* 2017:1–7. <https://doi.org/10.1155/2017/8314740>
- Spencer B, Wang G (2017) Statistical image reconstruction for shortened dynamic PET using a dual kernel method. In: *IEEE nuclear science symposium and medical imaging conference (NSS/MIC)*, pp 1–3. <https://doi.org/10.1109/NSSMIC.2017.8532801>
- Šprem J, de Vos BD, de Jong PA, Viergever MA, Išgum I (2017) Classification of coronary artery calcifications according to motion artifacts in chest CT using a convolutional neural network. In: *Medical imaging: image processing, SPIE*, vol 10133, p 101330R. <https://doi.org/10.1117/12.2253669>
- Sun C, Guo S, Zhang H, Li J, Chen M, Ma S, Jin L, Liu X, Li X, Qian X (2017a) Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs. *Artif Intell Med* 83(SI):58–66. <https://doi.org/10.1016/j.artmed.2017.03.008>
- Sun C, Guo S, Zhang H, Li J, Ma S, Li X (2017b) Liver lesion segmentation in CT images with MK-FCN. In: *IEEE 2nd advanced information technology, electronic and automation control conference (IAEAC)*, pp 1794–1798. <https://doi.org/10.1109/IAEAC.2017.8054322>

- Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 35(5):1299–1312. <https://doi.org/10.1109/TMI.2016.2535302>
- Tang Y, Cai J, Lu L, Harrison AP, Yan K, Xiao J, Yang L, Summers RM (2018) CT image enhancement using stacked generative adversarial networks and transfer learning for lesion segmentation improvement. In: *Machine learning in medical imaging*. Springer, pp 46–54
- Teramoto A, Fujita H, Yamamuro O, Tamaki T (2016) Automated detection of pulmonary nodules in PET/CT images: ensemble false-positive reduction using a convolutional neural network technique. *Med Phys* 43(6Part1):2821–2827. <https://doi.org/10.1118/1.4948498>
- Tomita N, Cheung YY, Hassanpour S (2018) Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med* 98:8–15. <https://doi.org/10.1016/j.compbiomed.2018.05.011>
- Townsend DW (2004) Physical principles and technology of clinical PET imaging. *Ann Acad Med* 33(2):133–45
- Townsend DW (2008) Positron emission tomography/computed tomography. *Semin Nucl Med* 38(3):152–166. <https://doi.org/10.1053/j.semnuclmed.2008.01.003>
- Trullo R, Petitjean C, Nie D, Shen D, Ruan S (2017a) Fully automated esophagus segmentation with a hierarchical deep learning approach. In: *IEEE international conference on signal and image processing applications (ICSIPA)*, pp 503–506. <https://doi.org/10.1109/ICSIPA.2017.8120664>
- Trullo R, Petitjean C, Nie D, Shen D, Ruan S (2017b) Joint segmentation of multiple thoracic organs in CT images with two collaborative deep architectures. In: *International workshop on deep learning in medical image analysis (DLMIA)*, vol 10553, pp 21–29. [https://doi.org/10.1007/978-3-319-67558-9\\_3](https://doi.org/10.1007/978-3-319-67558-9_3)
- Trullo R, Petitjean C, Dubray B, Ruan S (2019) Multiorgan segmentation using distance-aware adversarial networks. *J Med Imaging* 6(01):1. <https://doi.org/10.1117/1.JMI.6.1.014001>
- Tseng HH, Luo Y, Cui S, Chien JT, Ten Haken RK, Naqa IE (2017) Deep reinforcement learning for automated radiation adaptation in lung cancer. *Med Phys* 44(12):6690–6705. <https://doi.org/10.1002/mp.12625>
- Umehara K, Näppi JJ, Hironaka T, Regge D, Ishida T, Yoshida H (2017) Deep ensemble learning of virtual endoluminal views for polyp detection in CT colonography. In: *Medical imaging: computer-aided diagnosis, SPIE*, vol 10134, p 101340G. <https://doi.org/10.1117/12.2255606>
- van Tulder G, de Bruijne M (2016) Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines. *IEEE Trans Med Imaging* 35(5):1262–1272. <https://doi.org/10.1109/TMI.2016.2526687>
- Vaquero JJ, Kinahan P (2015) Positron emission tomography: current challenges and opportunities for technological advances in clinical and preclinical imaging systems. *Annu Rev Biomed Eng* 17(1):385–414. <https://doi.org/10.1146/annurev-bioeng-071114-040723>
- Vial A, Stirling D, Field M, Ros M, Ritz C, Carolan M, Holloway L, Miller AA (2018) The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. *Transl Cancer Res* 7(3):803–816. <https://doi.org/10.21037/tcr.2018.05.02>
- Vivanti R, Szeskin A, Lev-Cohain N, Sosna J, Joskowicz L (2017) Automatic detection of new tumors and tumor burden evaluation in longitudinal liver CT scan studies. *Int J Comput Assist Radiol Surg* 12(11):1945–1957. <https://doi.org/10.1007/s11548-017-1660-z>
- Wadsak W, Mitterhauser M (2010) Basics and principles of radiopharmaceuticals for PET/CT. *Eur J Radiol* 73(3):461–469. <https://doi.org/10.1016/j.ejrad.2009.12.022>
- Wang Y (2018) Convolutional neural network based malignancy detection of pulmonary nodule on computer tomography. Master of science, University of Saskatchewan
- Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W, Liu W, Yu L (2017a) Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Res* 7(1):11. <https://doi.org/10.1186/s13550-017-0260-9>
- Wang Y, Qiu Y, Thai T, Moore K, Liu H, Zheng B (2017b) A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. *Comput Methods Programs Biomed* 144(405):97–104. <https://doi.org/10.1016/j.cmpb.2017.03.017>
- Wang Y, Liao Y, Zhang Y, He J, Li S, Bian Z, Zhang H, Gao Y, Meng D, Zuo W, Zeng D, Ma J (2018a) Iterative quality enhancement via residual-artifact learning networks for low-dose CT. *Phys Med Biol* 63(21):215004. <https://doi.org/10.1088/1361-6560/aae511>
- Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, Wu X, Zhou J, Shen D, Zhou L (2018b) 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *NeuroImage* 174(8):550–562. <https://doi.org/10.1016/j.neuroimage.2018.03.045>

- Wang Z, Li J, Enoch M (2019) Removing ring artifacts in CBCT images via generative adversarial networks with unidirectional relative total variation loss. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-018-04007-6>
- Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I (2016) Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal* 34:123–136. <https://doi.org/10.1016/j.media.2016.04.004>
- Wolterink JM, Leiner T, Viergever MA, Išgum I (2017) Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging* 36(12):2536–2545. <https://doi.org/10.1109/TMI.2017.2708987>
- Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q, Shen D (2017) Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing* 267:406–416. <https://doi.org/10.1016/j.neucom.2017.06.048>
- Xie Y, Zhang J, Xia Y, Fulham M, Zhang Y (2018) Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf Fus* 42:102–110. <https://doi.org/10.1016/j.inffus.2017.10.005>
- Xie H, Yang D, Sun N, Chen Z, Zhang Y (2019a) Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognit* 85:109–119. <https://doi.org/10.1016/j.patcog.2018.07.031>
- Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M, Cai W (2019b) Knowledge-based collaborative deep learning for benign–malignant lung nodule classification on chest CT. *IEEE Trans Med Imaging* 38(4):991–1004. <https://doi.org/10.1109/TMI.2018.2876510>
- Xu J, Liu H (2017) Segmentation of pulmonary CT Image by using convolutional neural network based on membership function. In: *IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC)*, pp 198–203. <https://doi.org/10.1109/CSE-EUC.2017.42>
- Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, Piraud M, Buck A, Shi K, Menze BH (2018) Automated whole-body bone lesion detection for multiple myeloma on 68 Ga-pentixafor PET/CT imaging using deep learning methods. *Contrast Media Mol Imaging* 2018:1–11. <https://doi.org/10.1155/2018/2391925>
- Xu M, Qi S, Yue Y, Teng Y, Xu L, Yao Y, Qian W (2019) Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset. *BioMed Eng OnLine* 18(1):2. <https://doi.org/10.1186/s12938-018-0619-9>
- Xue Y, Chen S, Qin J, Liu Y, Huang B, Chen H (2017) Application of deep learning in automated analysis of molecular images in cancer: a survey. *Contrast Media Mol Imaging* 2017:1–10. <https://doi.org/10.1155/2017/9512370>
- Yan Z, Zhan Y, Peng Z, Liao S, Shinagawa Y, Zhang S, Metaxas DN, Zhou XS (2016) Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition. *IEEE Trans Med Imaging* 35(5):1332–1343. <https://doi.org/10.1109/TMI.2016.2524985>
- Yang H, Yu H, Wang G (2016) Deep learning for the classification of lung nodules. *CoRR arXiv:1611.06651*
- Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Zhang Y, Sun L, Wang G (2018) Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Trans Med Imaging* 37(6):1348–1357. <https://doi.org/10.1109/TMI.2018.2827462>
- Yi X, Babyn P (2018) Sharpness-aware low-dose CT denoising using conditional generative adversarial network. *J Digit Imaging* 31(5):655–669. <https://doi.org/10.1007/s10278-018-0056-0>
- You C, Yang Q, Shan H, Gjesteby L, Li G, Ju S, Zhang Z, Zhao Z, Zhang Y, Cong W, Wang G (2018) Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. *IEEE Access* 6:41839–41855. <https://doi.org/10.1109/ACCESS.2018.2858196>
- Ypsilantis PP, Siddique M, Sohn HM, Davies A, Cook G, Goh V, Montana G (2015) Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. *PLoS ONE* 10(9):e0137036. <https://doi.org/10.1371/journal.pone.0137036>
- Yuan Y (2017) Hierarchical convolutional–deconvolutional neural networks for automatic liver and tumor segmentation. *CoRR* i:3–6
- Yuan J, Liu X, Hou F, Qin H, Hao A (2018) Hybrid-feature-guided lung nodule type classification on CT images. *Comput Graph* 70:288–299. <https://doi.org/10.1016/j.cag.2017.07.020>
- Zahedinasab R, Mohseni H (2018) Enhancement of CT brain images classification based on deep learning network with adaptive activation functions. In: *IEEE 8th international conference on computer and knowledge engineering (ICCKE)*, pp 182–187. <https://doi.org/10.1109/ICCKE.2018.8566362>
- Zhang Y, Yu H (2018) Convolutional neural network based metal artifact reduction in X-ray computed tomography. *IEEE Trans Med Imaging* 37(6):1370–1381. <https://doi.org/10.1109/TMI.2018.2823083>
- Zhang Y, He Z, Zhong C, Zhang Y, Shi Z (2017) Fully convolutional neural network with post-processing methods for automatic liver segmentation from CT. In: *IEEE Chinese automation congress (CAC)*, pp 3864–3869. <https://doi.org/10.1109/CAC.2017.8243454>

- Zhao F, Xie X (2013) An overview of interactive medical image segmentation. *Ann BMVA* 2013(7):1–22
- Zhao X, Li L, Lu W, Tan S (2018a) Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. *Phys Med Biol* 64(1):015011. <https://doi.org/10.1088/1361-6560/aaf44b>
- Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W (2018b) Agile convolutional neural network for pulmonary nodule classification using CT images. *Int J Comput Assist Radiol Surg* 13(4):585–595. <https://doi.org/10.1007/s11548-017-1696-0>
- Zhao L, Lu Z, Jiang J, Zhou Y, Wu Y, Feng Q (2019) Automatic nasopharyngeal carcinoma segmentation using fully convolutional networks with auxiliary paths on dual-modality PET-CT images. *J Digit Imaging* 32(3):462–470. <https://doi.org/10.1007/s10278-018-00173-0>
- Zheng Y, Liu D, Georgescu B, Nguyen H, Comaniciu D (2015) 3D deep learning for efficient and robust landmark detection in volumetric data. In: International conference on medical image computing and computer-assisted intervention (MICCAI). Springer, pp 565–572. [https://doi.org/10.1007/978-3-319-24553-9\\_69](https://doi.org/10.1007/978-3-319-24553-9_69)
- Zhong Z, Kim Y, Zhou L, Plichta K, Allen B, Buatti J, Wu X (2018) 3D fully convolutional networks for co-segmentation of tumors on PET-CT images. In: IEEE 15th international symposium on biomedical imaging (ISBI), pp 228–231. <https://doi.org/10.1109/ISBI.2018.8363561>
- Zhou X, Ito T, Takayama R, Wang S, Hara T, Fujita H (2016) Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting. In: Deep learning and data labeling for medical applications. Lecture notes in computer science, vol 10008. Springer, pp 111–120. [https://doi.org/10.1007/978-3-319-46976-8\\_12](https://doi.org/10.1007/978-3-319-46976-8_12)
- Zhou X, Takayama R, Wang S, Hara T, Fujita H (2017a) Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med Phys* 44(10):5221–5233. <https://doi.org/10.1002/mp.12480>
- Zhou X, Takayama R, Wang S, Zhou X, Hara T, Fujita H (2017b) Automated segmentation of 3D anatomical structures on CT images by using a deep convolutional network based on end-to-end learning approach. In: Medical imaging: image processing, SPIE, vol 10133, p 1013324. <https://doi.org/10.1117/12.2254201>
- Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L (2019) Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol* 58(2):257–264. <https://doi.org/10.1080/0284186X.2018.1529421>
- Zreik M, van Hamersvelt RW, Wolterink JM, Leiner T, Viergever MA, Isgum I (2018) A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. *IEEE Trans Med Imaging*. <https://doi.org/10.1109/TMI.2018.2883807>

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