A Comprehensive Review of Computer-Aided Diagnosis of Pulmonary Nodules Based on Computed Tomography Scans

WENMING CAO[1,2], **(Senior Member, IEEE), RUI WU**[1], **GUITAO CAO**[3,4], **(Member, IEEE), AND ZHIHAI HE**[2], **(Fellow, IEEE)**

[1]Guangdong Multimedia Information Service Engineering Technology Research Center, Shenzhen University, Shenzhen 518060, China
[2]Video Processing and Communication Laboratory, Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211, USA
[3]Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai 200062, China
[4]MOE Research Center for Software/Hardware Co-Design Engineering, East China Normal University, Shanghai 200062, China

Corresponding author: Guitao Cao (gtcao@sei.ecnu.edu.cn)

ABSTRACT Lung cancer is one of the malignant tumor diseases with the fastest increase in morbidity and mortality, which poses a great threat to human health. Low-Dose Computed Tomography (LDCT) screening has been proved as a practical technique for improving the accuracy of pulmonary nodule detection and classification at early cancer diagnosis, which helps to reduce mortality. Therefore, with the explosive growth of CT data, it is of great clinical significance to exploit an effective Computer-Aided Diagnosis (CAD) system for radiologists on automatic nodule analysis. In this article, a comprehensive review of the application and development of CAD systems is presented. The experimental benchmarks for nodule analysis are first described and summarized, covering public datasets of lung CT scans, commonly used evaluation metrics and various medical competitions. We then introduce the main structure of a CAD system and present some efficient methodologies. For the extensive use of Convolutional Neural Network (CNN) based methods in pulmonary nodule investigations recently, we summarized the advantages of CNNs over traditional image processing methods. Besides, we mainly select the CAD systems developed by state-of-the-art CNNs with excellent performance and analyze their objectives, algorithms as well as results. Finally, research trends, existing challenges, and future directions in this field are discussed.

INDEX TERMS CAD, lung cancer, pulmonary nodules detection, classification, CT scans, CNN.

I. INTRODUCTION

Over the past decades, cancer treatment has been the critical focus of medical research on human health all around the world. According to the 2020 cancer report released by World Health Organization, cancer is the second leading cause of death globally, with an estimated 9.6 million people deaths in 2018, accounting for one in six deaths [1]. The global cancer burden is heavy and growing. Among various types of cancers, lung cancer is the most frequently diagnosed cancer with the highest rate of incidence and mortality, as shown in Fig. 1. In the United States, lung cancer death rate continuously declined from 2008 to 2017, yet it still caused more deaths in 2017 than breast, prostate, colorectal, and brain

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang.

cancers combined [2]. Many risk factors, such as outdoor air pollution and the prevalence of tobacco use, contribute to most of the deaths and disease from lung cancer. However, lung cancer treatment is becoming more and more unaffordable, and healthcare systems are struggling to provide new cancer cures for people. Therefore, lung cancer interventions, including primary prevention, screening, and early diagnosis, remain a top priority and are more meaningful for people to reduce financial and psychological barriers [3]. Pulmonary nodule analysis is one of the effective cancer prevention interventions, consisting of detection step and classification step.

Generally, pulmonary nodules are characterized as a round opacity or irregular lung lesions with diameters from approximately 3 mm to 30 mm, which can be solitary or multiple. They are complex in number (single or multiple), size
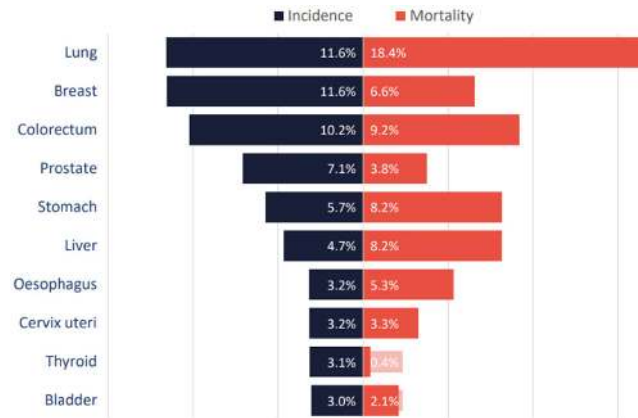
FIGURE 1. Most common cancer cases in 2018 all around the world.

(diameters$\leq$8mm or diameters>8mm), shape (round, polygonal or irregular), margins (smooth, lobulated or spiculate), location (well-circumscribed, juxta-pleural or juxta-vascular) and density (solid, subsolid or ground glass nodule). Some types of nodules are shown in Fig. 2. These nodules are common, and most of them are benign. In many cases, lesions of pulmonary nodules are often asymptomatic and can increase the difficulty of diagnosis. However, studies show that nodules with large size (diameters>8mm), subsolid, spiculate, lobulated characteristics are more likely to be malignant [16], [17]. From [10], the 5-year survival rate is only 10-15% for patients diagnosed with lung cancer, while the rate for those with early-stage cancerous lesion completely resected increased to 65-80%. It is challenging but essential to determine whether the nodule is malignant in an early stage.
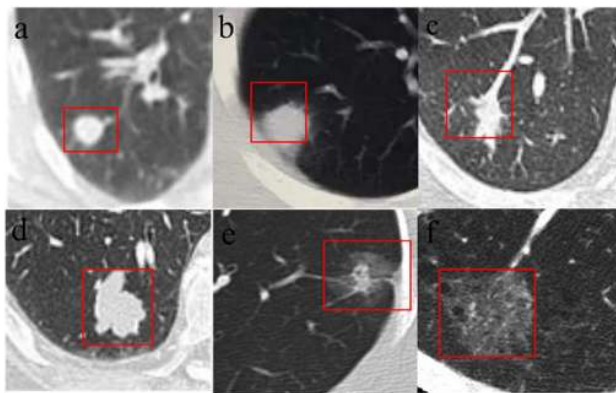

FIGURE 2. Example of multiple pulmonary nodules: (a) smooth and well-circumscribed nodule, (b) juxta-pleural nodule, (c) juxta-vascular nodule, (d) lobulated and solid nodule, (e) subsolid nodule, and (f) ground glass nodule.

Low-dose computed tomography (LDCT) Screening dramatically reduces mortality from lung cancer [8]. To be specific, the National Lung Screening Trial (NLST), which is the largest trial to date, demonstrated that a 20.0% decline in mortality rate among participants at high risk for lung cancer compared with chest radiography screening. More recently, the Multicentric Italian Lung Detection Trial showed a 39%

reduction in mortality compared with no early intervention [2], [9]. LDCT screening detects more pulmonary nodules and has lower radiation damage, which is helpful for radiologists to diagnose early-stage lung cancers and make treatment plans. Thus, the open datasets, evaluation metrics, and developing algorithms that we select in this article, are all related to LDCT acquisition techniques.

With the popularization of LDCT screening techniques, an enormous increase in CT scans burdens radiologists heavily. Manual analysis in massive CT scans is becoming a very tedious and time-consuming task. Therefore, with the aim of lessening the radiologist's workload, an efficient Computer-Aided Diagnosis (CAD) system is necessary to facilitate the process of automatically analyzing large amounts of CT scans. In recent years, CAD systems have been widely adopted in addressing various diseases [3]. For its high validity and reliability in clinical diagnosis, the global CAD market estimated to reach 2.7 billion dollars by 2025, expanding at a compound annual growth rate of 11.6% over the forecast period [15]. Specifically, a conventional CAD system can be divided into a detection system (CADe) and a diagnostic system (CADx). CADe aims to locate the interest regions of the lung CT scans to detect abnormal lesions. CADx is designed to assist radiologists or clinicians in determining the type, malignancy of the anomalies. In general, a CAD system for lung cancer lies the emphasis on detection and classification of pulmonary nodules, consisting of three stages: (1) preprocessing, (2) nodule detection, including candidate nodule detection and false positive reduction, (3) nodule classification. Preprocessing is mainly conducted to reduce noise, segment the Regions Of Interest (ROI) in the lung for narrowing the search range of pulmonary nodules, and normalize the data. In the nodule detection stage, candidate nodules should be detected as many as possible, which often results in high sensitivity and low accuracy. Then the false positive reduction step should be performed to catch precise nodules marks. Finally, the classification stage aims to predict the probability of nodule malignancy [4]. Practical algorithms of three stages of CAD development will be covered in more detail in section III.

Numerous published works have been applied to improve the performance of CAD systems for pulmonary nodule analysis. In the case of insufficient resources and dataset, researchers usually use traditional machine learning methods such as multiple gray-level thresholding, linear discriminant analysis, distance transformation and Support Vector Machine (SVM) for quick investigation on lung nodules [18]–[23]. However, using Deep Learning approaches in various medical imaging tasks has been a mainstream trend in the past ten years. Deep Neural Networks (DNNs), especially Convolutional Neural Networks (CNNs) have repeatedly been shown the outstanding performance in many open computer vision competitions, including ImageNet challenges and Microsoft Common Objects in Context (MS COCO) challenges. For the high adaptability of CNNs, many CNN-based models, such as U-Net, Faster Region CNN

(Faster R-CNN), Mask-RCNN, and Retina-Net [24]–[29], have been widely applied on nodule detection and classification tasks, increasing accuracy and robustness of CAD systems.

This article aims to provide a detailed overview of CAD systems for pulmonary nodule detection and classification, which can be used as a study guide for researchers. Compared with the previous surveys [6], [7], this review not only illustrates the applications, experimental benchmarks, and construction constitutes of CAD systems but also emphasizes the introduction of various systems developed based on state-of-the-art CNNs. Accurately, we completely summarize the robust and effective algorithms for pulmonary nodule analysis, which have been validated on public or large datasets with excellent performance. After eliminating similar papers, we only detailly analyze those that have shown the best performance in lung cancer diagnosis to avoid duplicate content from previous reviews. We mainly focus on works published in 2019 and 2020 with the latest advancements, yet a small part of relevant works proposed before 2019 are also included. Please refer to the Appendix section for details of the literature collection procedure.

The remainder of this article is organized as follows. Firstly, the experimental benchmarks on pulmonary nodule detection and classification, including public datasets of lung CT scans, widely used evaluation methods, and related competitions, are introduced. Secondly, the complete structure of CAD systems, as well as some efficient algorithms of each component, are explicitly presented. Thirdly, workflows of CNN-based algorithms and traditional image processing methods are presented, and the advantages of CNN-based algorithms are summarized. Fourthly, the CAD systems, which are developed using state-of-the-art CNNs with excellent performance, are analyzed. Finally, research trends, current challenges, and prospective directions of CAD system development for pulmonary nodule analysis are discussed.

II. EXPERIMENTAL BENCHMARKS

For the development of effective CAD systems, there are three experimental benchmarks for researchers to focus on: datasets, evaluation metrics, and large-scale competitions for lung cancer diagnosis. Training pulmonary nodule detection and classification models require a large volume of lung CT scans, thus the acquisition of public datasets is extremely vital. To fairly validate the performance of various algorithms, reliable evaluation metrics are necessary. In addition, large-scale competitions always provide up-to-date CAD models, which are trained based on unified datasets and evaluation standards.

A. PUBLIC DATASETS OF LUNG CT SCANS

1) NLST
Dataset of the National Lung Screening Trial was collected on lung cancer diagnoses and deaths through Dec. 31, 2009, with 54000 participants enrolled from 2002 and 2004 randomly

assigning to an LDCT group and a chest radiography group. Approximately 200,000 image series from over 75,000 CT exams are available, which include data on participant characteristics, screening exam results, diagnostic procedures, lung cancer, and mortality [31].

2) VIA/I-ELCAP
The International Early Lung Cancer Action Program database was made for the performance evaluation of diverse CAD systems by the ELCAP and Vision and Image Analysis research groups. 50 LDCT scans with a slice thickness of 1.25mm, nodule location information, and nodules types are provided. Particularly, the nodule sizes of this database are relatively small [32].

3) NELSON
Nederlands-Leuvens Longkanker Screenings Onderzoek trial was designed to investigate the benefits of LDCT screening on lung cancer mortality. Data of 15822 participants were collected since 2003. Datasets were original from images of the lung with a thickness of 1 mm and reconstructed at an overlap interval of 0.7 mm [34].

4) LIDC-IDRI
The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) is the largest publicly available reference database for lung nodules. This database contains 1018 CT scans and associated XML files of two-phase image annotations from four experiences radiologists. Additionally, annotation consists of nodule characteristics, types, and position [36].

5) OTHERS
Datasets of Non-Small Cell Lung Cancer (NSCLC)-Radiomics, ACRIN-NSCLC-FDG-PET, LungCT-Diagnosis, and QIN LUNG CT are relatively less-used for improving the sensitivity of CAD system in contrast experiments. However, researchers can use these datasets for model robustness and generalization testing. These datasets can be downloaded from The Cancer Imaging Archive (TCIA) [42]. The information of all datasets is summarized in Table 1.

B. EVALUATION METRICS
The commonly used evaluation metrics of pulmonary nodule detection and classification are listed below:

1) TPR, FPR, ACCURACY
Ture Positive Rate (TPR), False Positive Rate (FPR), and accuracy are the primary methods to quantitatively analyze the experimental results in nodule detection and classification. Four variables are used in calculating these three parameters: Ture Positive (TP), Ture Negative (TN), False Positive (FP), False Negative (FN). Specifically, TPR, which is also called sensitivity or recall, shows the relationship between TPs and the total number of actual positive samples, using Eq. (1). FPR is calculated as the ratio between FPs and

TABLE 1. Diverse public datasets with a large number of lung CT scans.

Database	Release Year	Patients Number	Sample Number	Modality	Data Access	Collection Status	Annotation
VIA/I-ELCAP [32]	2003	N/A	50	CT	Available	Complete	Nodule position and type
NELSON [35]	2003	15822	N/A	CT	N/A	Complete	N/A
NLST [31]	2009	54000	75000+	CT	Available	Complete	Participant characteristics and diagnostic information
LIDC-IDRI [37]	2011	N/A	1018	CT, CR, DX	Available	Complete	Nodule characteristics, type and position
NSCLC-Radiomics [38]	2014	422	1265	CT, RTSTRUCT, SEG	Available	Ongoing	Participant characteristics and diagnostic information
ACRIN-NSCLC-FDG-PET [39]	2013	242	3377	PT, CT, MR, CR, DX, SC, NM	Available	Complete	Clinical data
LungCT-Diagnosis [40]	2015	61	61	CT	Available	Complete	Image position
QIN LUNG CT [41]	2015	47	47	CT	Available	Complete	N/A

Note: Not Available (N/A), Computed Radiography (CR), Digital radiography (DX), Radiotherapy Structure set (RTSTRUCT), DICOM segmentation (SEG), Positron Emission Tomography-Computed Tomography (PT/PET-CT), Magnetic Resonance (MR), Secondary Capture (SC), Nuclear Medicine (NM)

the total number of actual negative samples, using Eq. (2). Besides, accuracy describes the closeness of the positive predicted values to actual samples, using Eq. (3).

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

2) ROC, FROC
Receiver Operating Characteristic curve (ROC) and Free-Response Operating Characteristic curve (FROC) methods are generally used to analyze the performance of a CADe system, which visually estimate the pros and cons of various comparative algorithms. A ROC curve indicates the relationship between sensitivity and FPR, with FPR on the X-axis and sensitivity on the Y-axis. An FROC curve is similar to a ROC curve except that the number of FPs per scan replaces the FPR. Particularly, the FROC method is often used in most competitions related to medical image processing.

3) CPM
Competition Performance Metric (CPM) refers to the average sensitivity at seven predefined FPRs: 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per scan, which is defined as the final score of FROC curve. CPM was first proposed in Automatic Nodule Detection 2009 (ANODE09) [50] and became the evaluation criteria of most pulmonary nodule detection competitions. CPM is calculated using Eq. (4). Visually, Fig. 3 shows the
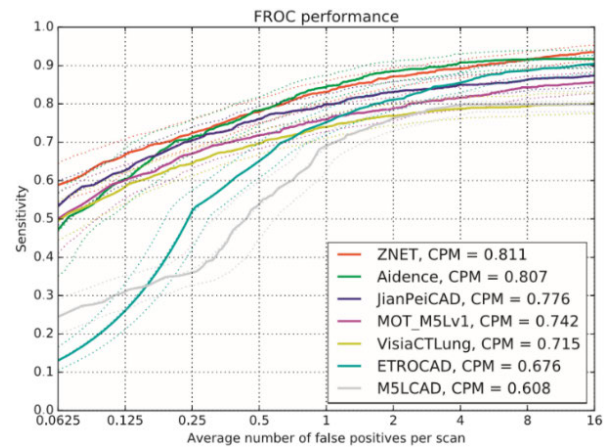

FIGURE 3. The FORC curves and CPMs of various systems.

FORC curves and CPMs of various CAD systems from [30].

$$CPM = \frac{1}{7} \sum_{i=FPs} s\,(i),\ FPs = \{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8\} \tag{4}$$

where i represents the number of FP per scan at seven predefined FPR levels, s means the sensitivity of a CAD system, which can be referred to Eq. (1).

4) AUC, LOGLOSS
The Area Under the Curve (AUC) of the ROC curve and LogLoss are usually applied to evaluate the algorithms' ability to predict malignancy and classify lung cancer. The larger the AUC, the higher the classification accuracy of a CAD system. On the contrary, the lower the LogLoss, the better

TABLE 2. Grand challenges on pulmonary nodules detection and classification.

Challenge	Year	Sample Number	Annotation	Data Access	Constituent Parts	Evaluation Method	Best Performance
ANODE [43]	2009	5 example and 50 test scans	Nodule type and position	Available	NDET	FROC, **CPM**	0.632
LUNA [44]	2016	888 CT scans with 1186 nodules, ST≅2.5mm	Nodule size and position	Available	NDET	FROC, **CPM**, CV	0.951
					FPRED	FROC, **CPM**, CV	0.968
TIANCHI [45]	2017	1000 CT scans for S1, 2000 CT scans for S2, ST≅2mm	Nodule size and position	N/A	NDET	FROC, **CPM**, CV	S1:0.815 S2:0.732
DSB [46]	2017	1594 CT scans for S1, 506 CT scans for final test	Cancer ground truth of S1 data	N/A	NDCF	**LogLoss**	0.3998
ISBI-LNMP [47]	2018	200 CT scans for 100 subjects	Nodule position and time interval	N/A	NDCF	AUC	0.913
LNDb [48]-[48]	2019	294 CT scans	Nodule characteristics, type and position	Available	Fleischner CF	**Fleiss-Cohen weighted Cohen's kappa**, CV	0.5795
					NDET	**ACPM**, CV	0.5332

Note: (1) Nodule Detection (NDET), False Positive Reduction (FPRED), Nodule Classification (NDCF), Nodule Segmentation (NDSG), Slice Thickness (ST), Nodule Texture Characterization (ND TETCRT), Average Competition Performance Metric at two levels (ACPM), Stage 1/2 (S1/2);
(2) The bold words in Evaluation Method column indicate the method of obtaining the final score, corresponding to Best Performance column

performance of a CAD system. The definition of LogLoss is:

$$\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i\log\left(\hat{y}_i\right) + (1 - y_i)\log\left(1 - \hat{y}_i\right)\right] \quad (5)$$

where n is the number of patients in the test set, $\hat{y}_i$ is the predicted probability of the image belonging to a patient with cancer, y_i is one if the diagnosis is cancer, zero otherwise.

5) CROSS-VALIDATION
Cross-validation (CV) is a statistical method used to evaluate the performance of machine learning models. It is largely used in the phase of training predictive models because it can avoid overfitting and improve model generalization when there is not enough available data. Specifically, splitting the dataset into a test set and a training set in different ways is the critical part. CV methods contain K-Fold CV, Leave One Group Out, and Holdout method.

C. LARGE-SCALE COMPETITIONS RELATED TO PULMONARY NODULE
With the development of CAD, various competitions related to lung nodules have created a vital learning platform for researchers. Most of the medical image processing competitions provide datasets, effective evaluation methods, and reference algorithms, which are suitable learning sources for researchers and developers. In this article, we choose large-scale competitions that are related to pulmonary nodule detection and classification. Each of the competitions contains over 50 even to 3000 participants joined. The selected competitions consist of ANODE09, Lung Nodule Analysis 2016 (LUNA16), TianChi Medical AI competition (TIANCHI) on the intelligent diagnosis of pulmonary nodules, Kaggle Data Science Bowl 2017 (DSB), International Symposium on Biomedical Imaging-Lung Nodule

Malignancy Prediction based on sequential CT scans (ISBI-LNMP) and Lung Nodule Database-Automatic Lung Cancer Patient Management (LNDb) [66]. Table 2 shows detailed information of each competition, including sample data, constituent parts, evaluation criteria, best results, etc.

III. STRUCTURE OF A CAD SYSTEM
For decades, to increase the efficiency of lung cancer diagnosis, numerous researches have been done. Potentially, CAD systems can take advantage of thin cross-section images and serve as a second interpreter for radiologists for pulmonary nodule identification. There are a variety of CAD systems designed with different structures, and the main structure includes three components: (1) Preprocessing, (2) Nodule Detection, including candidate nodule detection and false positive reduction, (3) Nodule Classification. The whole procedure on how a CAD system works is shown in Fig. 4. Performance varies significantly among CAD systems due to the CT input, different characteristics of nodules, and especially the diversity of algorithms. To improve both sensitivity and specificity, most of the studies are focused on false positive reduction and nodule classification while using the same datasets.

In this section, we completely summarize the components of a CAD system and present practical algorithms that have been proved the effectiveness on public datasets.

A. PREPROCESSING
Preprocessing is a significant first stage of lung CT image analysis, because much irrelevant information existing in raw images that reduces working efficiency and diagnostic accuracy of a CAD system. The main lung volume, which is the ROI, is the core searching space while conducting nodule detection. Therefore, removing distracting components such
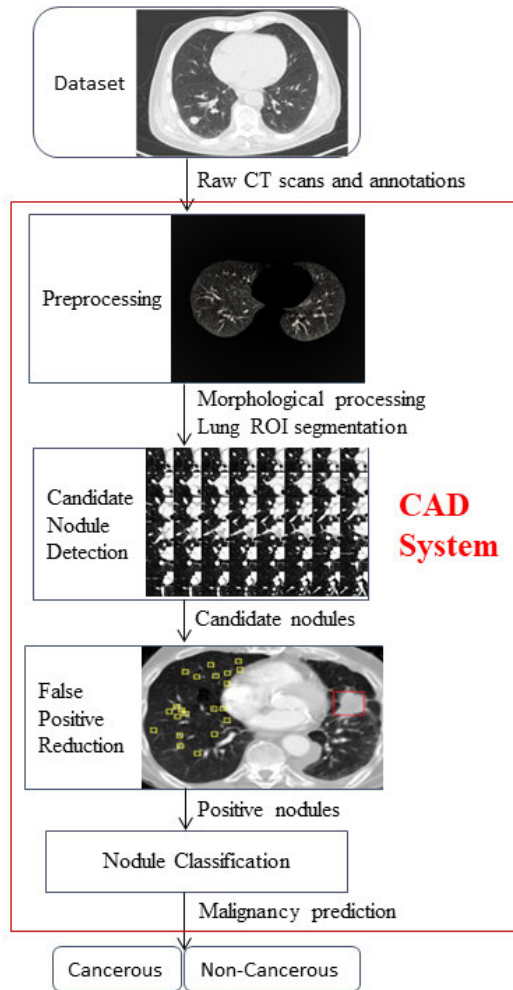
FIGURE 4. The complete workflow of a CAD system.

as chest tissues and image artifacts, recovering or enhancing useful information are the key goals in this stage. It is illustrated that adapting lung segmentation procedures, as a preprocessing step in a CAD system, can prevent approximately 5% - 17% missing of detected nodules [33]. Specifically, the contrast of Hounsfield Unit (HU) value between the lung and surrounding tissue forms the basis of most segmentation methods, and these methods can be divided into rule-based approaches and data-based approaches [4].

Generally, thresholding, component analysis, region growing, morphological operations, and filtering [19], [22], [26], [51], [52], [62], [67], [68], [96], [109] are often used as rule-based approaches in preprocessing medical images. Thresholding and component analysis are the most effective and quick ways to approximately separate lung volume from distracting components. Then lung volume can be found by restricting size and location. Also, using region growing can identify lung volume from trachea and bronchi. After determining lung volume, morphological operations, such as erosion and dilation, can be performed to obtain nonporous, smooth-bound lung. Various filters, such as Gaussian smooth

filter and arithmetic mean filter, are recommended to reduce noise or enhance the image quality. These methods can be combined in various ways to achieve different segmentation effects. Han *et al.* [19] used thresholding to simply extract the chest volume, then applied a two-class high-level vector quantization algorithm for classifying voxels, followed by a linear Karhunen-Lo'eve transformation of the local intensity vectors. Besides, the principal component analysis technique was performed to optimize the vector space. Liao *et al.* [62] firstly adopted a Gaussian filter, intensity and distance thresholding operations to extract the mask of lung and rule out other tissues, next performed morphological operations including convex hull computing and dilation, to optimize mask extraction. Fig. 5 shows a complete preprocessing procedure using rule-based approaches.
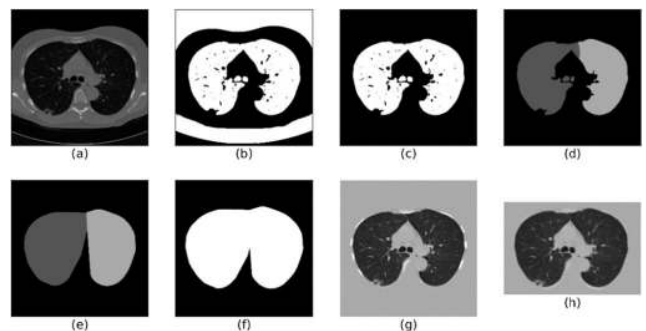

FIGURE 5. The procedures of preprocessing. (a) Convert the CT image into HU format, (b) Filtering then thresholding image for binarization, (c) select the connected domain of lung volume, (d) separate the left and right lungs, (e) calculate the convex hull on each of lung side. (f) dilate and merge the two masks, (g) multiply the raw image with the mask, fill the masked area with tissue luminance, and transform the image to UINT8, (h) crop the image into a specific size.

Furthermore, data-based approaches with better applicability can be used after rule-based operations by training learnable models to refine lung segmentation. Soliman *et al.* [63] initially identified the background of all 3D chest scans by region growing and component analysis. They then proposed a joint 3D Markov-Gibbs Random Field (MGRF) framework, which integrated two appearance sub-models and an adaptive shape prior sub-model to segment both normal and pathological lungs. Besides, deep learning techniques, particularly CNNs can be applied for medical image segmentation when there is sufficient hardware support. Existing advanced CNNs such as U-net [64], Mask-RCNN [65], and hybrid CNNs [69] can also be used for automatic lung segmentation. Alom *et al.* [69] proposed a Recurrent CNN (RU-Net) and a Recurrent Residual CNN (R2U-Net) for medical image segmentation, both of which were designed based on U-Net models. The proposed CNNs were evaluated on the LUNA16 dataset to segment lung region and achieved an accuracy of 99.18%.

In fact, rule-based approaches can also reach similar segmentation performance by manually adjusting parameters as data-based approaches. However, data-based approaches cost more time to train a learnable model, and it will be more

computationally expensive than using rule-based approaches in optimizing CAD systems. Rule-base approaches are more convenient options for researchers to preprocess lung CT images.

B. NODULE DETECTION

Nodule detection involves nodule registration and filtering, called candidate nodule detection and false positive reduction, respectively. For decades, various nodule detection techniques have been proposed because of the complexity of nodule texture, size, shape and location, etc. We roughly divide diverse techniques into two common categories: traditional methods and DNN-based methods. The traditional methods, mainly consisting of classical image processing methods and machine learning classifiers, determine nodules and eliminate FPs by maximizing the matching rate between the feature profiles and suspicious area under the premise of manual-defined specific features. On the other hand, DNN-based methods, especially CNN-based methods, refer to black-box operations, which extract implicit features and tuning system performance automatically. Many practical algorithms of nodule detection are proposed based on both the traditional and DNN-based methods [77].

In this part, we briefly introduce useful algorithms which are developed for candidate nodule detection and false positive reduction. The effective algorithms proposed in 2019 and 2020 are selected and summarized in Table 3.

1) CANDIDATE NODULE DETECTION

In this stage, the main goal of CAD is to generate as many candidates as possible without considering the specificity but sensitivity. The more the pulmonary nodules are detected, the higher the survival rate of patients. Candidate Nodule Detection (CNDET) is a procedure to identify suspicious lesions and provide predicted position and probability of candidates.

For decades, several traditional methods such as thresholding, region growing, clustering, distance transformations, and morphological operations [19]–[23], [70]–[73], [83], [100] have been widely used based on hand-crafted features for roughly recognizing candidate nodules. El-Regaily *et al.* [73] first applied rule-based approaches, including contrast-enhancing, region growing, rolling-ball algorithm and morphological operations, to extract lung parenchyma as well as preserve nodules attached to the lung wall. They then used 3D region growing, Euclidean distance transform and 2D thresholding to capture candidates from depth maps. However, these classical image processing methods are developed according to pixel intensity and low-level representatives of images. Additional filtering or geometric features computing methods are also needed to optimize candidate nodule generation [22], [74].

With the popularization of deep learning, more and more detection algorithms are proposed based on DNN techniques. So far, a large number of DNN-based methods, specifically CNN-based methods, are applied to generate candidate

nodules because they can capture both low-level features and abstract high-level features, which greatly improve the detection sensitivity. The commonly used network structures for nodule detection mainly comprise simple CNN, U-Net, Feature Pyramid Network (FPN), Region Proposal Network (RPN), Residual Network (ResNet), and Retina-Net. Nearly all of the detection algorithms are variants of these networks, [26]–[28], [51], [53]–[57], [61], [75]–[79]. Part of them are Hybrid networks, which combine multiple structures in cascade mode [24], [28], [62], [68], [90], [99]. Wang *et al.* [75] proposed a nodule-size-adaptive model, which is similar to Faster R-CNN, to locate candidates with bounding boxes. Azad *et al.* [68] developed a Bi-directional ConvLSTM U-Net with Densely connected convolutions (BCDU-Net), which use different ways of concatenation to take full advantages of multiple feature maps for lung nodule recognition and segmentation. Besides, some fusion networks are also explored using multi-stream structures in order to integrate the power of different networks [25], [65], [80]–[82]. Liu *et al.* [80] exploited three identical 3D ResUNets to generate 3D Gaussian blob nodules, then fine-tuned the network by adding 3D RPN heads resulting in higher sensitivity on large nodules.

2) FALSE POSITIVE REDUCTION

After the CNDET stage, there are still many FPs decreasing the efficiency of nodule diagnosis. Excessive FPs will lead to over-diagnosis, over-treatment, waste of medical resources. Therefore, it is essential to increase the accuracy of nodule detection by reducing FPs. False Positive Reduction (FPRED) refers to classify true nodules from extracted candidates, which is equivalent to a binary classification task. There are also many works focusing on FPRED.

In the FPRED stage, a variety of features based on intensity, morphology, or texture should be extracted from candidate nodule images and fed to classifiers to determine nodule and non-nodule candidates. For traditional methods, several machine learning classifiers are commonly applied to recognize true nodules, for example, SVM, k-Nearest Neighborhood classifiers, linear discriminant classifiers, and various boosting classifiers [19]–[22], [83], [84]. Naqi *et al.* [22] combined geometric texture and Histogram of Oriented Gradient reduced by Principle Component Analysis (HOGPCA) features into a hybrid feature vector, then fed the extracted vector to k-Nearest Neighbourhood, Naive Bayesian, SVM, and AdaBoost to reduce FPs.

In recent years, massive DNN-based methods, specifically various CNNs, have been proposed to enhance classification performance. According to the difference of network structures, we can categorize these networks into advanced off-the-shelf CNNs [24], [51], [54], [67], [76], [78]–[80], [94], [100], [105], [108], and multi-stream heterogeneous CNNs [53], [58], [72], [75], [85], [86]–[88]. Liu *et al.* [76] developed a High Sensitivity and Specificity (HS^2) network, which was built by two convolution layers and three fully connected layers, to track the appearance

TABLE 3. Practical works for candidate nodule detection and false positive reduction.

Author	Year	Method	Dataset	Best Performance
Li et al. [55]	2019	CNDET: 3D CNN using intersection over union self-normalization and maxout methods	LUNA16	CPM: 0.912
El-Regaily et al. [73]	2019	CNDET: 2D thresholding, 3D region growing, Euclidean distance transform FPRED: rule-based classifier, followed by multi-view 2D CNN	LIDC-IDRI	Sensitivity: 0.853
Wang et al. [75]	2019	CNDET: CNN consisting of a feature extraction network, an RPN and a region classification network FPRED: CNN-based classifier built with two inception-v4 neural networks	LUNA16	CPM: 0.903
Xiao et al. [85]	2019	FPRED: Multi-scale heterogeneous 3D CNN	LUNA16	CPM: 0.874
Zuo et al. [58]	2019	FPRED: Multi-Resolution features integrated 2D CNN	LUNA16	CPM: 0.762
Liu et al. [76]	2019	CNDET: 3D FPN with pre-trained ResNet-18 model FPRED: HS^2 Network	LUNA16	CPM:0.957
			SPIE-AAPM [102]	CPM:0.899
			LungTIME [103]	CPM:0.889
			HMS Lung Cancer [104]	CPM:0.871
Cheng et al. [62]	2019	CNDET: 2D modified U-Net with residual blocks FPRED: 3D multi-classification networks including VGG, Inception, and DenseNet	TIANCHI	Sensitivity: 0.73
Tang et al. [78]	2019	CNDET: 3D RPN adapted from Faster R-CNN FPRED: 3D CNN	TIANCHI	CPM:71.16
Kim et al. [86]	2019	FPRED: Multi-scale Gradual Integration CNN	LUNA16	CPM: 0.942
Zhou et al. [105]	2019	FPRED: 2/3D Models Genesis with encoder-decoder architecture	LUNA16	CPM:0.971 AUC:0.982
Shi et al. [108]	2019	FPRED: pre-trained VGG-16 and an SVM classifier	LIDC-IDRI	Sensitivity: 0.872 at 0.39 FPs/scan
Liu et al. [80]	2020	CNDET: 3D ResU-Net and 3D RPN FPRED: 3D Dense U-Net with simple classification head	NLST	CPM:0.879
Wang et al. [99]	2020	CNDET: 3D modified FPN for FocalMix method using SSL	LUNA16: Labelled data NLST: Unlabelled data	CPM:0.907
Zheng et al. [81]	2020	CNDET: 2D fusion U-Nets based on MIP FPRED: 3D fusion VGG-nets	LIDC-IDRI	Sensitivity: 0.942 at 2 FPs/scan
Zuo et al. [88]	2020	FPRED: Multi-branch 3D CNN	LUNA16	CPM: 0.83
Zheng et al. [79]	2020	CNDET: 2D U-net++ FPRED: 3-D multi-scale Dense CNN	LIDC-IDRI	CPM: 0.955
Ozdemir et al. [51]	2020	CNDET: 3D segmentation network based on V-net architecture FPRED: 3D scoring network	LUNA16	CPM: 0.921
Masood et al. [54]	2020	CNDET: Multi-PRN inspired by modified VGG-16 FPRED: 3D CNN using modified ResNet-101 basic layout	LUNA16	CPM: 0.946
			LIDC-IDRI	Sensitivity: 0.974
			ANODE09	Sensitivity: 0.988

Note: Semi-Supervised Learning (SSL), Maximum Intensity Projection (MIP). The method column shows the algorithms used in the CNDET and the FPRED steps.

changes in continuous CT slices of each nodule candidate on Location History Images (LHI). Cheng *et al.* [53] applied a 3D multi-classification network composed of VGG, Inception, and Dense networks to eliminate FPs, which achieved a sensitivity of 73% on the TIANCHI dataset.

Besides, optimization strategies such as data augmentation, positive samples balancing via focal loss function, and Non-Maximum Suppression (NMS) operations [80]can be used to improve performance for better classification. For example, At Liu *et al.* [80] applied Projected Gradient Descent to generate three types of adversarial samples, then trained a 3D Dense U-Net with the extracted candidates and adversarial samples, resulting in a 5.33% improvement on CPM.

TABLE 4. Useful works for nodule classification.

Author	Year	Method	Dataset	Best Performance
Al-Shabi et al. [11]	2019	Gated-Dilated network, using multiple dilated convolution operations	LIDC-LDRI	AUC: 0.9514
Xie et al. [56]	2019	3D multi-view knowledge-based collaborative DNN containing three pre-trained ResNet-50 networks	LIDC-IDRI	AUC: 0.957
Nasrullah et al. [90]	2019	CMixNet and gradient boosting machine	LIDC-IDRI	Sensitivity: 0.94
Liao et al. [62]	2019	Leaky Noise-or Network using data augmentation and alternate training scheme	LUNA16 and DSB	LogLoss:0.399 AUC:0.87
Balagurunathan et al. [91]	2019	Optimal linear classifier	NLST	AUC: 0.85
Al-Shabi et al. [92]	2019	Deep Local-Global Networks containing residual blocks and non-local blocks	LIDC-IDRI	Accuracy: 0.885 AUC: 0.956
Gao et al. [96]	2019	DLSTM (convolutional version) using a TEM	NLST	AUC:0.826
Chen et al. [98]	2019	Pre-train Med3D models based on various ResNets, then transfer Med3D models to initialize networks of the classification task	LIDC-IDRI	Accuracy: 0.919
Ardila et al. [65]	2019	Cancer risk prediction model based on 3D Inception blocks	NLST	AUC: 0.944
Hussein et al. [106]	2019	Pre-trained 3D CNN, K-mean clustering and proportion-SVM, using both supervised and unsupervised learning algorithms	LIDC-IDRI	Accuracy: 0.786
Yang et al. [94]	2020	Self-attention transformer equipped based on 3D DenseNet, using MIL method	LIDC-IDRI	AUC: 0.932
Harsono et al. [29]	2020	Integrated modified pre-trained Inflated 3D ConvNet with FPN	LIDC-IDRI	AUC: 0.818
Apostolopoulos et al. [93]	2020	Dual deep solitary pulmonary nodules network modified from DPN	LIDC-IDRI	Accuracy: 0.93
Liu et al. [61]	2020	Multi-task CNN based on residual blocks with margin ranking loss	LIDC-IDRI	AUC: 0.9797

Note: Distanced Long Short-Term Memory (DLSTM), Temporal Emphasis Model (TEM), Multiple Instance Learning (MIL). The method column shows the algorithms used in the classification step.

C. NODULE CLASSIFICATION

Nodule classification is the final step of CAD systems. Most of the CAD systems are designed for predicting malignancy of nodules and determining whether a nodule is cancerous, but some are designed for nodule type classification [21], [95]. In this article, we mainly focus on lung cancer diagnosis. The age, sex, pack-years smoked, and smoking status of patients can cause nodule lesions. Since cancerous nodules tend to have large size (diameters>8mm), and uneven surface with spiculate, lobulated characteristics, measurements of nodule size and representations of nodule appearance are the most important research direction to estimate the malignant probability.

A number of classification techniques are applied in this stage: (1) machine learning classifiers, such as SVM, k-Nearest Neighborhood classifiers, Bayesian classifiers, boosting classifiers, optimal linear classifiers [59], [89], [91]; (2) advanced off-the-shelf CNNs [11], [51], [55], [62], [92], [93]; (3) CNNs integrated with machine learning classifiers [28], [90], [97]; (4) multi-stream heterogeneous CNNs [56], [61], [67], [106]; (5) CNNs trained with transfer learning algorithms [29], [56], [61], [60], [67], [94], [98], [106]. As can be seen from the above, there are multi-stream hybrid CNNs trained with transfer learning techniques. For instance, Xie et al. [56] performed a 3D multi-view

knowledge-based collaborative (MV-KBC) deep model, which was built with three pre-trained ResNet-50 networks, to extract multiple features from nine planes and diagnose the malignancy of nodules. Table 4 shows the useful works proposed in 2019 and 2020 on nodule classification.

IV. CNNS VS. TRADITIONAL METHODS
A. WORKFLOW

Rapid improvements in computing power, as well as an increase in the amount of available data, enable the extensive use of DNN-based methods in medical image processing. Particularly, CNNs are responsible for the tremendous influence in the field of CAD development, and significantly improve the accuracy in nodule detection and classification tasks. CNNs are designed to discover the underlying relationship between images and automatically extract the most descriptive features, mainly in an end-to-end manner. CNNs are typically built by three types of layers (convolution layers, pooling layers and fully connected layers.) and activation functions. The convolution and pooling layers perform feature extraction while the fully connected layers map the extracted features to the final output. And each fully connected layer is followed by a specific activation function such as sigmoid, softmax and ReLU. The activation functions are selected according to different data and
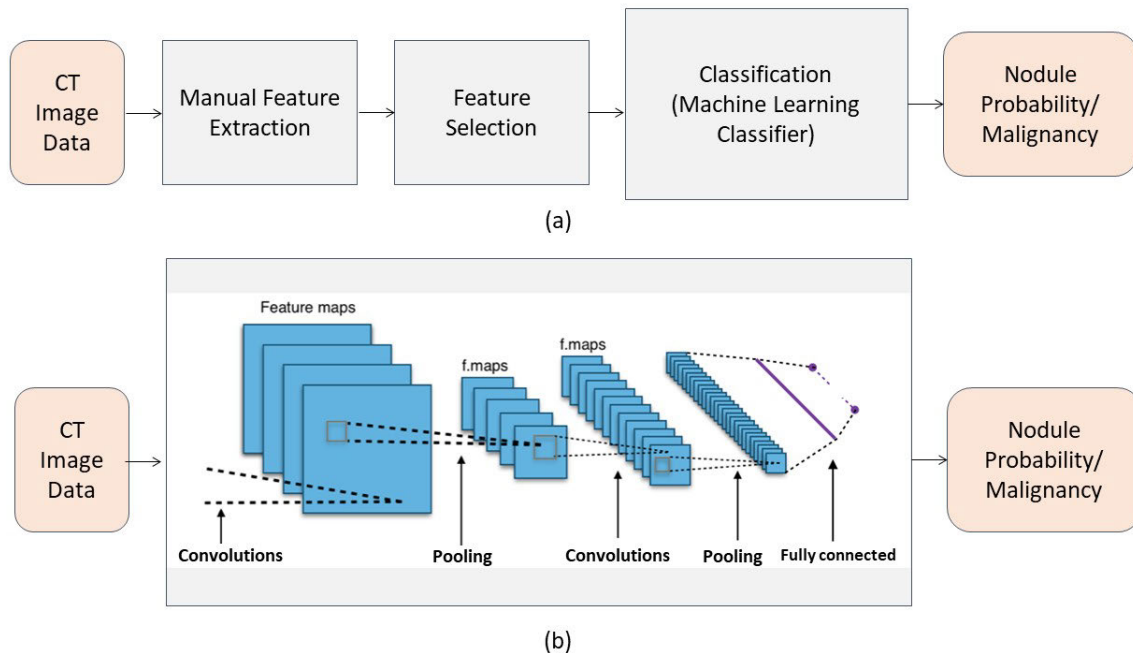
FIGURE 6. (a) Tradition methods workflow VS. (b) Typical CNNs workflow.

classification tasks. However, traditional methods are to applied several computer vision techniques for image processing. For feature extraction, it is necessary to determine hand-crafted features and manually select the important ones in each given image, which heavily depend on the subjective judgment of researchers. In the following step, machine learning classifiers need to be performed for nodule classification. The workflows using typical CNNs and traditional methods for nodule diagnosis are described in Fig. 6.

B. ADVANTAGES OF CNNs
1) BETTER PERFORMANCE
Unlike traditional methods, CNNs utilize the shared convolution kernel to discover potential patterns in different image categories, which are beneficial for processing not only single-dimensional but also high-dimensional data. This characteristic of CNNs enables the exploration of high-level semantic information from substantial medical images, leading to better performance in image detection and classification tasks.

2) STRONG FLEXIBILITY
CNNs provide strong flexibility and adaptivity on various datasets. Since CNNs are mathematical models with approximation functions, any dataset that can be quantified can be used to retrain CNN-based models for both regression and classification problems. In contrast to traditional methods, classical image processing methods and machine learning classifiers tend to be domain-specific [107].

3) AUTOMATED AND EFFICIENT
CNNs are designed based on black-box operations, which are able to automatically extract descriptive and salient features

corresponding to each target object. However, the difficulty of traditional methods is to manually define and select specific features according to different image tasks. As the number of image categories increases, feature extraction becomes more time-consuming and energy-draining.

V. STATE-OF-THE-ART
Part of the algorithms in this article involve a small extension to previously published works, or are just tested with different datasets. Therefore, only the CAD systems with the best performance or developed with state-of-the-art CNNs will be detailedly introduced in this section.

A. MULTI-STREAM FRAMEWORKS
Multi-stream frameworks refer to those developed using multi-scale, multi-resolution, multi-views input data, or those designed with multiple networks. Applying multi-stream frameworks can take advantage of different types of features for better identifying malignant nodules.

1) MULTI-SCALE GRADUAL INTEGRATION CADE
The Multi-scale Gradual Integration CNN (MGI-CNN) CADe system, which was designed specifically for false positive reduction, was adapted by Kim *et al.* [86]. MGI-CNN was proposed to extract morphological and contextual features from multi-scale input data. The MGI-CNN consisted of two main components: Gradual Feature Extraction (GFE) and Multi-Stream Feature Integration (MSFI). The candidate nodule patches were first extracted from thoracic CT scans at three different scales: $40 \times 40 \times 26$, $30 \times 30 \times 10$, $20 \times 20 \times 6$. The extracted patches were fed into 'zoom-in' and 'zoom-out' GFE networks next. Then the multiple-stream features were fused by MSFI to integrate contextual

information hierarchically. The algorithm was evaluated on the LUNA16 dataset and obtained CPM scores of 0.908(V1) and 0.942(V2). Note that the source code is available at: https://github.com/ ku-milab/MGICNN.

2) MAXIMUM INTENSITY PROJECTION-BASED CADE

This CADe system was developed by Zheng *et al.* [81] using Maximum Intensity Projection (MIP)-based CNNs. The framework of the proposed CADe included four fusion 2D CNNs based on U-Net and a 3D CNN based on VGG-Net. MIP images were the superposition of maximum grey values at each coordinate from a set of consecutive slices. Specifically, ruled-base methods, covering thresholding, component analysis, and binary morphology operations, were first used to segment lung volume. Then the MIP images with different slab thicknesses were generated as input for 2D CNNs to detect candidates from four streams. All the fused candidates were finally fed into a 3D CNN to reduce FPs. In the CNDET stage, the MIP-based CADe system was trained on the LUNA16 dataset and achieved a sensitivity of 95.36% with 20.4 FPs/scan. In the FPRED stage, the system was performed on the LIDC-IDRI dataset and obtained a sensitivity of 94.19% with 2 FPs/scan.

3) CLOUD-BASED AUTOMATED CADE

Masood *et al.* [54] developed a CADe system, which was integrated cloud computing provided by virtual machines and software given by a 3D CNN model. In the preprocessing stage, Median Intensity Projection (MeIP) was applied to generate MeIP images. Then traditional methods were used to extract multi-scale, multi-angle, and multi-view patches from MeIP images. In the CNDET stage, a multi-Region Proposal Network (mRPN) architecture was built based on a modified VGG-16 backbone to detect candidate nodules from extracted patches. They chose seven levels of anchor sizes to generate nodules of diverse malignant levels: 4×4, 8×8, 12×12, 16×16, 20×20, 26×26, and 32×32. In the following stage, a 3D CNN using a modified ResNet-10 basic layout was performed to reduce FPs. The Cloud-based automated CADe system was trained and validated on LUNA16, ANODE09, LIDC-IDR datasets, achieving sensitivities of 0.988 at 1.97 FPs/scan, 0.976 at 2.3 FPs/scan and 0.988 at 1.97 FPs/scan, respectively.

4) END-TO-END CADX

Ardila *et al.* [65] proposed an end-to-end 3D CNN framework that trained with patients' current and prior CT volumes for lung cancer diagnosis. The architecture of the CADx system consisted of four sectors: (1) Lung segmentation model: a Mask-RCNN was trained on the LUNA16 dataset to produce segmentation masks of lung CT scans. (2) Cancer ROI detection model: a modified 3D Retina-Net was pre-trained on the LIDC dataset and fine-tuned on the NLST dataset to generate nodule-like ROIs. (3) Full-volume model: a 3D inflated Inception V1 was trained on the 1.5 mm^3 voxel size CT volumes for cancer prediction, fine-tuning from a checkpoint

trained on ImageNet. (4) Cancer risk prediction model: a 3D Inception was applied to extract features from the output of the (2)(3) models for final malignancy prediction, exploiting both nodule-level local information and global context from the entire CT volume. The CADx system achieved the best performance on NLST data (AUC=0.944).

5) MULTI-VIEW CADX

Xie *et al.* [56] adopted a U-Net to segment the lung nodules on a slice-by-slice basis. Next, a 3D multi-view knowledge-based collaborative (MV-KBC) deep model consisting of nine KBC sub-models was trained to learn multiple characteristics from nine specific views (sagittal, coronal, axial and six diagonal planes) on each 3D nodule image. Each sub-models comprised three pre-trained ResNet-50 networks, which were applied to extract the overall appearance, heterogeneity in voxel values, heterogeneity in shapes patches of nodules on each plane. Besides, a penalty loss function was introduced to balance the number of positive and negative samples. The proposed model was tested on the LIDC-IDRI dataset for nodule classification, resulting in an accuracy of 91.6% and the AUC of 95.7%.

B. TRANSFER LEARNING

Transfer learning indicates an algorithm that stores knowledge gained while solving a specific task and can be applied to another relevant task. Initialize or fine-tune models using pre-trained CNNs can improve the efficiency and accuracy of CAD systems to some extent.

1) MULTI-PLANAR CADE

Zheng *et al.* [79] proposed a 2D U-Net++ and a 3D Multi-Scale Dense CNN to develop the CADe system for small nodule identification. This CADe system was trained using images from the axial plane, the coronal plane, and the sagittal plane with transfer learning algorithms. For image preprocessing, the multi-planar lung volumes were segmented using the same ways as [81]. For candidate nodule generation, the multi-planar slices and MIP slices served as the input of the U-Net++ model, which was adapted from efficient-net pre-trained on ImageNet. The U-Net++ model extracted features from both small and large receptive fields and detected candidate nodules on each plane using bounding boxes. Next, a 3D Multi-Scale Dense CNN, consisting of thirty-two basic blocks, five transition blocks, and a classifier block, was applied to exclude suspicious candidates. The Multi-planar CADe system was trained and validated on the LIDC-IDRI dataset. It reached a sensitivity of 0.981 when identifying candidate nodule, and obtained a CPM score of 0.955 after reducing FPs.

2) MED3D CADX

The Med3D was developed by Chen *et al.* [98] for 3D medical image segmentation and classification using transfer learning algorithms. The procedure of the Med3D CAD contained three steps: (1) The 3DSeg-8 dataset with diverse scans

regions, target organs, and pathologies were collected and normalized. (2) The Med3D, which consisted of a shared encoder adapted from ResNets and eight simple decoder branches, was trained to extract specific features. (3) Transfer the pre-trained Med3D to different medical tasks. In the lung cancer diagnosis task, they changed the encoder part of Med3D to a ResNet classifier for feature extraction, then added an average pooling layer and a full-connected layer with (1,1,1) kernel size for malignant nodule classification. The system evaluation was conducted on the LIDC dataset with an accuracy of 0.919 and higher network convergence speed. Note that all pre-trained models and source codes are provided at: https://github.com/Tencent/MedicalNet.

C. SEMI/UN-SUPERVISED BASED CADe
Semi-supervised learning methods can be used to automatically extract features from a very limited number of labeled data and a large number of unlabeled data. In contrast, unsupervised learning methods are used on unlabeled data. Semi/un-supervised learning algorithms can increase the accuracy of lung cancer diagnosis in the case of insufficient training data.

Wang *et al.* [99] proposed a FocalMix method, which was to take advantage of the latest semi-supervised learning (SSL) algorithms for 3D medical image processing. FocalMix method mainly included three optimization strategies to improve the effectiveness of lung nodule analysis: soft-target focal loss, anchor-level target prediction model, and MixUp augmentation. The labeled images and unlabeled images were utilized as input data. Firstly, the training anchors in labeled images were assigned according to the annotated boxes, while the unlabeled ones were extracted by the target prediction model. Besides, the target prediction model was designed using both traditional methods and CNN-based methods in SSL manner, covering image transformations, morphological operations, and a 3D variant of FPN. After that, two levels of MixUp augmentation, which were image-level Mixup and Object-level Mixup, were applied to each input batch. Additionally, the soft-target focal loss was used on unlabeled data to train the model. They evaluated the FocalMix method on the LUNA16 dataset and NLST dataset, resulting in a CMP score of 0.907. It was proved that the proposed method outperformed the fully supervised baseline, and was easy to transfer to other modern SSL frameworks.

D. SELF-SUPERVISED LEARNING
Self-supervised learning is an approach that makes full use of unlabeled data to generate the needed information for supervised feature learning.

1) MODELS GENESIS CADE
The Models Genesis (MG) CADe system was designed by Zhou *et al.* [105]. The MG consisted of various 2/3D source models, which were trained from unlabeled images using a unified self-supervised learning method. The proposed

system was built in an encoder-decoder architecture and was applied in different imaging tasks. The MG consolidated four novel transformations: (1) non-linear, (2) local-shuffling, (3) out-painting, and (4) in-painting, to recover anatomical patterns. The MG was trained from diverse perspectives (appearance, texture, context, etc.) by unifying all tasks into a single image restoration task via transformation operations. Models fine-tuned from MG outperformed models learned from scratch and any 2D models in five target tasks, including both image segmentation and classification. For nodule and non-nodule classification tasks, the MG CADe system was tested on the LUNA16 dataset and achieved the best performance (AUC=0.982, CPM=0.971). Note that all pre-trained models are available at: https://github.com/MrGiovanni/ModelsGenesis

2) HIGH SENSITIVITY AND SPECIFICITY CADX
Liu *et al.* [76] developed a CADx using self-supervised learning. The whole framework of this CADx system comprised a 3D FPN and a High Sensitivity and Specificity (HS^2) network. In the preprocessing stage, traditional methods such as Gaussian filters were applied to segment lung region. After that, a 3D FPN with a self-supervised pre-trained ResNet-18 was adopted for candidate nodule recognition by using multi-scale features to improve the resolution of nodules and parallel top-down path to transfer high-level semantic features to supplement low-level features. For false positive reduction, the HS^2 network consisting of two convolution layers and three fully connected layers was performed to track the appearance changes of each candidate from continuous CT slices on Location History Images (LHI). The proposed system achieved CPM scores of 0.957, 0.899, 0.889, 0.871 on LUNA16, SPIE-AAPM, LungTIME, and HMS Lung Cancer datasets, respectively.

E. MULTI-TASK CADx
Multi-task learning is a method that solves multiple learning tasks at the same time and finds out the commonalities and differences across tasks. Compared with the single-task training model, multi-task learning can improve the learning efficiency and accuracy of the specific task model.

Liu *et al.* [61] proposed a multi-task 2D CNN with Margin Ranking loss (MTMR-Net) to construct the CADx system for nodule analysis. The MTMR-Net consisted of two 2D CNNs in Siamese network architecture for nodule benign-malignant classification task and attribute score regression task. Besides, a margin ranking loss was employed to further classify ambiguous nodules, which improved the discriminating capability of the network. Specifically, each 2D CNN was built with a feature extraction module, a classification module, and a regression module: (1) The feature extraction module was designed based on residual blocks and was trained using parameters from pre-trained ResNet-15. (2) The classification module contained one fully connected layer followed by a cross-entropy loss for the final benign-malignant classification. (3) The regression module

comprised two fully connected layers followed by mean square error loss for the final attribute score prediction (eight attribute scores: internal structure, calcification, sphericity, margin, spiculation, lobulation, and texture). They evaluated the proposed CADx on the LIDC-IDRI dataset, obtaining an accuracy of 0.935, a sensitivity of 0.93, a specificity of 0.894, and an AUC of 0.9797, respectively. Note that source codes are available at: https://github.com/CaptainWilliam/MTMR-NET.

VI. DISCUSSION

From the investigation mentioned above of advanced CAD systems, it has been proved that remarkable progress has been achieved in automatic pulmonary nodule analysis. Particularly, various advanced CNN-based algorithms have been applied to improve the accuracy and sensitivity in nodule detection and classification tasks, thus significantly increasing the effectiveness of CAD systems for lung cancer diagnosis in an early stage. Despite more and more intelligent CAD systems appear with the popularization of CT scanning techniques and deep learning approaches, there are still problems existing. In this section, we analyze trends from the research works mentioned above and present some unsolved challenges and future directions in pulmonary nodule diagnosis.

A. RESEARCH TRENDS

As shown in Section III and IV, substantial CNN-based researches have been done for candidate nodule detection, false positive reduction, and nodule classification. It is observed that developers tend to perform CNN-based algorithms instead of traditional methods to design CAD systems in lung cancer identification with the rising of computer power [7].

From the state-of-the-art CAD systems listed in Section V, we categorize the corresponding developing strategies into five CNN groups: (1) multi-stream CNNs [54], [56], [65], [79], [81], [86], (2) CNNs with transfer learning algorithms [54], [56], [76], [79], [98], (3) CNNs with semi/un-supervised learning algorithms [99], (4) CNNs with self-supervised algorithms [76], [105], and (5) multi-task based CNNs [61]. It is shown that the ratio of using these five strategies in state-of-the-art CAD systems is 6:5:1:1, while the ratio of using these five strategies in practical publications proposed in 2019 and 2020 is 15:8:2:2:2 (see Table 3, 4). Obviously, multi-stream CNNs are more commonly adopted than other state-of-the-art strategies, possibly because the multi-stream framework can exploit comprehensive multi-modal features, including both low-level image features and high-level semantic features, leading to higher accuracy in nodule diagnosis. Furthermore, it is worth noting that some multi-stream CNNs with transfer learning algorithms [54], [56], [76], [79] obtain CPM and AUC over 94%, which outperform most of the other strategies. The reason for its strong stability and great performance is that the extracted common features from different disease patterns, as well

as multi-modal features, may have significant benefits for detection and classification tasks.

B. EXISTING CHALLENGES AND FUTURE DIRECTIONS
1) LACK OF A LARGE AND HIGH-QUALITY LABELED DATASETS

As is well known, a large amount of high-quality labeled data is crucial to train an effective deep learning model for medical image analysis. However, the existing public datasets of lung CT scans are not labeled in an organized manner, which results in messy annotated information among different datasets. Thus collecting mass lung CT data with accurate labels remains a big challenge. On the one hand, privacy issues could be the biggest obstacle for collecting individual lung CT scans, and some hospital regulations and national policies also involve personal information protection. On the other hand, it takes much time for radiologists to annotate the medical images while works to non-expertise would lead to misclassification.

To alleviate the dataset scarcity problem, data augmentation strategies such as cropping, rotation, flipping, or scaling of image patches and relevant labels can be applied to increase the number and diversity of available training samples. Besides, Generative Adversarial Networks (GAN) can also be adopted to synthesize adversarial images as additional data [110]. When there are sufficient raw CT scans with lack of labels, advanced off-the-shelf CNNs can be trained on much or all of the unlabeled data using semi/un-supervised as well as self-supervised learning methods, which will reach a better performance than supervised learning methods [76], [99], [105], [106]. Using transfer learning algorithms to pre-train 3D CNNs on other large-scale datasets, such as ImageNet, will improve the accuracy in nodule detection and classification tasks in the case of insufficient datasets [60], [79].

2) BAD INTERPRETABILITY OF DIAGNOSTIC RESULT

CNN-based models are trained in the black-box procedure, which can automatically identify and classify pulmonary nodules, yet provide no explanation of pathogenesis. The interpretability of models is essential for radiologists to find out the exact cause of the disease. Only detective results or diagnosis scores do not significantly help radiologists make a final diagnosis and draw up an accurate treatment plan. Thus the CNN-based models, which can discover the relationship between input data and diagnostic results as well as determine which features of nodules are responsible for the malignancy, deserve attention.

To improve the interpretability of CAD system, an inference model based on the Bayesian network has been constructed using the Markov Chain Monte Carlo (MCMC) method, which can estimate the conditional probability of each feature [111]. Furthermore, the cause and effect inference problem could be divided into a feature prediction task and a benign-malignant classification task [25]. The casual

relationship between predicted feature scores and diagnostic results can be observed. For instance, a multi-task CNN system with margin ranking loss was proposed for nodule attribute score prediction and cancer diagnosis [61].

3) LACK OF CONTINUOUS LEARNING ABILITY

An effective CAD system for pulmonary nodule diagnosis is usually required to assist radiologists in making clinical decisions correctly when facing unexpected samples. Therefore, the continuous learning ability of a CAD system for new medical image samples identification is vital. However, present CAD systems are mostly developed using trained models and applied of actual use, which means that they can only perform well in statistic environments instead of a dynamic environment. These systems can not correctly recognize some unique samples that have not been trained, probably resulting in wrong diagnosis [77]. It is of great benefit to construct a CAD system with continuous learning ability to support real-time changing situations.

One possible direction for building continuously learning-based systems is to design a new CNN framework with cloud computing techniques. With the use of cloud computing techniques, diagnosis records can be sent to cloud storage to update training datasets so that the proposed CNN can be trained in a cloud back-end to adapt real-time changes continuously [54].

VII. CONCLUSION

In this article, we make a comprehensive review of pulmonary nodule detection and classification for CT images in developing a CAD system. The public datasets of lung CT scan, widely used evaluation methods, and related challenges of the pulmonary nodule are first introduced and summarized. Then we describe the detailed procedures on how a CAD system works and presented some practical algorithms of each processing stage. We next compare the differences between traditional methods and CNNs and summarize the advantage of CNNs. Besides, the CAD systems that are developed using state-of-the-art CNNs with excellent performance are filtered and analyzed. Finally, we make a discussion about research trends, existing challenges, and future directions of CAD system development.

It can be concluded from this review that CNN-based methods are dominant with better performance and outperform the traditional methods in both nodule detection and classification tasks. The exploration of multi-stream, semi/un-supervised, self-supervised, multi-task, and transfer learning methods, especially multi-stream and transfer learning approaches for improving the performance of CAD systems, deserve more attention. Note that we focus on clarifying the development of CAD systems and analyzing the effective CNN-based algorithms. It is believed that this review can provide a comprehensive reference for researchers and radiologists.

APPENDIX

Most of the literature pertinent to pulmonary nodules analysis in CAD development is collected from searches using the public online database search engines: IEEE Xplore, Science Citation Index Expanded, arXiv, SpringerLink, and ScienceDirect. The searching keywords are ''lung cancer'', ''nodule detection'', ''nodule classification'', ''false positive reduction'', ''CAD'', and ''computer-aided diagnosis''. Combine the keywords using ''OR'' and ''AND'' in different ways. Articles published in 2019 and 2020 with excellent performance are selected and filtered. Contributions from other sources can be identified from citations in the above publications.

REFERENCES

[1] *WHO Report on Cancer: Setting Priorities, Investing Wisely and Providing Care for All*, World Health Organization, Geneva, Switzerland, 2020.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, ''Cancer statistics, 2020,'' *Cancer J. Clinicians*, vol. 70, no. 1, pp. 7–30, Jan. 2020, doi: 10.3322/caac.21590.

[3] S. Cressman, ''The cost-effectiveness of high-risk lung cancer screening and drivers of program efficiency,'' *J. Thoracic Oncol.*, vol. 12, no. 8, pp. 1210–1222, Aug. 2017, doi: 10.1016/j.jtho.2017.04.021.

[4] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, ''Computer analysis of computed tomography scans of the lung: A survey,'' *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 385–405, Apr. 2006, doi: 10.1109/TMI.2005.862753.

[5] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. S. Tavares, ''Automatic 3D pulmonary nodule detection in CT images: A survey,'' *Comput. Methods Prog. Biomed.*, vol. 124, pp. 91–107, Feb. 2016, doi: 10.1016/j.cmpb.2015.10.006.

[6] G. Zhang, S. Jiang, Z. Yang, L. Gong, X. Ma, Z. Zhou, C. Bao, and Q. Liu, ''Automatic nodule detection for lung cancer in CT images: A review,'' *Comput. Biol. Med.*, vol. 103, pp. 287–300, Dec. 2018, doi: 10.1016/j.compbiomed.2018.10.033.

[7] P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao, and W. Qian, ''Detection and classification of pulmonary nodules using convolutional neural networks: A Survey,'' *IEEE ACCESS*, vol. 7, p. 17, 2019.

[8] T. National Lung Screening Trial Research Team, ''Reduced lung-cancer mortality with low-dose computed tomographic screening,'' *New England J. Med.*, vol. 365, no. 5, pp. 395–409, Aug. 2011, doi: 10.1056/NEJMoa1102873.

[9] H. J. de Koning, ''Reduced lung-cancer mortality with volume CT screening in a randomized trial,'' *New England J. Med.*, vol. 382, no. 6, pp. 503–513, Feb. 2020, doi: 10.1056/nejmoa1911793.

[10] M. V. Davis, ''The solitary pulmonary nodule,'' *JAMA*, vol. 190, no. 4, p. 89S-96S, Oct. 1964.

[11] M. Al-Shabi, H. K. Lee, and M. Tan, ''Gated-dilated networks for lung nodule classification in CT scans,'' *IEEE Access*, vol. 7, pp. 178827–178838, 2019, doi: 10.1109/ACCESS.2019.2958663.

[12] H. Abe, H. Macmahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz, and K. Doi, ''Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996–2001 RSNA scientific assemblies,'' *RadioGraphics*, vol. 23, no. 1, pp. 255–265, Jan. 2003, doi: 10.1148/rg.231025129.

[13] U. Pastorino, M. Silva, S. Sestini, F. Sabia, M. Boeri, A. Cantarutti, N. Sverzellati, G. Sozzi, G. Corrao, and A. Marchianá, ''Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: New confirmation of lung cancer screening efficacy,'' *Ann. Oncol.*, vol. 30, no. 7, pp. 1162–1169, Jul. 2019, doi: 10.1093/annonc/mdz117.

[14] H. Macmahon, R. Engelmann, F. M. Behlen, K. R. Hoffmann, T. Ishida, C. Roe, C. E. Metz, and K. Doi, ''Computer-aided diagnosis of pulmonary nodules: Results of a large-scale observer test,'' *Radiology*, vol. 213, no. 3, pp. 723–726, Dec. 1999, doi: 10.1148/radiology.213.3.r99dc27723.

[15] (2020). *Global Computer-Aided Detection Market 2.7 Billion by 2025*. [Online]. Available: https://www.ihealthcareanalyst.com/pre-screening-diagnostic-technology-adoption-computer-aided-detection-market/

[16] A. Snoeckx, P. Reyntiens, D. Desbuquoit, M. J. Spinhoven, P. E. Van Schil, J. P. van Meerbeeck, and P. M. Parizel, "Evaluation of the solitary pulmonary nodule: Size matters, but do not ignore the power of morphology," *Insights into Imag.*, vol. 9, no. 1, pp. 73–86, Feb. 2018.

[17] Q. Zhou, "China national guideline of classification, diagnosis and treatment for lung nodules," *Zhongguo Zhi*, vol. 19, no. 12, pp. 793–798, Dec. 2016, doi: 10.3779/j.issn.1009-3419.2016.12.12.

[18] P. P. R. Filho, A. C. D. S. Barros, G. L. B. Ramalho, C. R. Pereira, J. P. Papa, V. H. C. de Albuquerque, and J. M. R. S. Tavares, "Automated recognition of lung diseases in CT images based on the optimum-path forest classifier," *Neural Comput. Appl.*, vol. 31, no. 2, pp. 901–914, Feb. 2019, doi: 10.1007/s00521-017-3048-y.

[19] H. Han, L. Li, F. Han, B. Song, W. Moore, and Z. Liang, "Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 2, pp. 648–659, Mar. 2015, doi: 10.1109/JBHI.2014.2328870.

[20] M. Bergtholdt, R. Wiemker, and T. Klinder, "Pulmonary Nodule Detection Using a Cascaded SVM Classifier," in *Computer-Aided Diagnosis*, vol. 9785, G. D. Tourassi and S. G. Armato, Eds. Bellingham, WA, USA: SPIE, 2015.

[21] S. G. and J. P., "A fully-automated system for identification and classification of subsolid nodules in lung computed tomographic scans," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101586, doi: 10.1016/j.bspc.2019.101586.

[22] S. M. Naqi, M. Sharif, and I. U. Lali, "A 3D nodule candidate detection method supported by hybrid features to reduce false positives in lung nodule detection," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 26287–26311, Sep. 2019, doi: 10.1007/s11042-019-07819-3.

[23] M. S. Brown, P. Lo, J. G. Goldin, E. Barnoy, G. H. J. Kim, M. F. McNitt-Gray, and D. R. Aberle, "Toward clinically usable CAD for lung cancer screening with computed tomography," *Eur. Radiol.*, vol. 24, no. 11, pp. 2719–2728, Nov. 2014, doi: 10.1007/s00330-014-3329-0.

[24] H. Tang, D. R. Kim, and X. Xie, "Automated pulmonary nodule detection using 3D deep convolutional neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 523–526, doi: 10.1109/ISBI.2018.8363630.

[25] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, and J.-Z. Cheng, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images," *IEEE Trans. Med. Imag.*, vol. 36, no. 3, pp. 802–814, Mar. 2017, doi: 10.1109/TMI.2016.2629462.

[26] G. Li, W. Zhou, W. Chen, F. Sun, Y. Fu, F. Gong, and H. Zhang, "Study on the detection of pulmonary nodules in CT images based on deep learning," *IEEE Access*, vol. 8, pp. 67300–67309, 2020, doi: 10.1109/ACCESS.2020.2984381.

[27] W. Cao, R. Wu, and G. Cao, "A multi-modal Neural Network for Lung Nodule Detection with Low-Dose CT Images," *Basic Clin. Pharmacol. Toxicol.*, vol. 124, pp. 37–38, Apr. 2019.

[28] W. Zhu, C. Liu, W. Fan, and X. Xie, "DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 673–681, doi: 10.1109/WACV.2018.00079.

[29] I. W. Harsono, S. Liawatimena, and T. W. Cenggoro, "Lung nodule detection and classification from thorax CT-scan using RetinaNet with transfer learning," *J. King Saud Univ. Comput. Inf. Sci.*, Apr. 2020, pp. 1–8, doi: 10.1016/j.jksuci.2020.03.013.

[30] A. A. A. Setio, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017, doi: 10.1016/j.media.2017.06.015.

[31] *NLST Datasets*. Accessed: Aug. 15, 2020. [Online]. Available: https://cdas.cancer.gov/datasets/nlst/

[32] *VIA/I-ELCAP Datasets*. Accessed: Aug. 15, 2020. [Online]. Available: http://www.via.cornell.edu/databases/lungdb.html

[33] S. G. Armato and W. F. Sensakovic, "Automated lung segmentation for thoracic CT," *Academic Radiol.*, vol. 11, no. 9, pp. 1011–1021, Sep. 2004, doi: 10.1016/j.acra.2004.06.005.

[34] Y. Ru Zhao, X. Xie, H. J. de Koning, W. P. Mali, R. Vliegenthart, and M. Oudkerk, "NELSON lung cancer screening study," *Cancer Imag.*, vol. 11, no. 1A, pp. S79–S84, 2011, doi: 10.1102/1470-7330.2011.9020.

[35] *NELSON Datasets*. Accessed: Aug. 15, 2020. [Online]. Available: https://www.rug.nl/research/portal/datasets/nederlandsleuvens-longkanker-screenings-onderzoek-nelson(0c42f592-4f1a-4d99-be08-ed2916758bb2).html

[36] S. G. Armato, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Jan. 2011, doi: 10.1118/1.3528204.

[37] *LIDC-IDRI Datasets*. Accessed: Aug. 15, 2020. [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

[38] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. C. J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Commun.*, vol. 5, no. 1, Jun. 2014, Art. no. 1, doi: 10.1038/ncomms5006.

[39] M. Machtay, F. Duan, B. A. Siegel, B. S. Snyder, J. J. Gorelick, J. S. Reddin, R. Munden, D. W. Johnson, L. H. Wilf, A. DeNittis, N. Sherwin, K. H. Cho, S.-K. Kim, G. Videtic, D. R. Neumann, R. Komaki, H. Macapinlac, J. D. Bradley, and A. Alavi, "Prediction of survival by [18F]fluorodeoxyglucose positron emission tomography in patients with locally advanced non–small-cell lung cancer undergoing definitive chemoradiation therapy: Results of the ACRIN 6668/RTOG 0235 trial," *JCO*, vol. 31, no. 30, pp. 3823–3830, Sep. 2013, doi: 10.1200/JCO.2012.47.5947.

[40] O. Grove, A. E. Berglund, M. B. Schabath, H. J. W. L. Aerts, A. Dekker, H. Wang, E. R. Velazquez, P. Lambin, Y. Gu, Y. Balagurunathan, E. Eikman, R. A. Gatenby, S. Eschrich, and R. J. Gillies, "Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma," *PLOS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118261, doi: 10.1371/journal.pone.0118261.

[41] D. Goldgof, L. Hall, S. Hawkins, M. Schabath, O. Stringfield, A. Garcia, and R. Gillies. *QIN LUNG CT Datesets*. Accessed: Aug. 22, 2020. [Online]. Available: http://doi.org/10.7937/K9/TCIA.2015.NPGZYZBZ

[42] *The Cancer Imaging Archive (TCIA)*. Accessed: Aug. 15, 2020. [Online]. Available: https://www.cancerimagingarchive.net/

[43] *ANODE09—Grand Challenge*. Accessed: Aug. 15, 2020. [Online]. Available: https://anode09.grand-challenge.org/Results/

[44] *LUNA16—Grand Challenge*. Accessed: Aug. 15, 2020. [Online]. Available: https://luna16.grand-challenge.org/Results/

[45] *TIANCHI Medical AI Challenge: Intelligent Diagnosis of Pulmonary Nodules*. Accessed: Aug. 15, 2020. [Online]. Available: https://tianchi.aliyun.com/competition/entrance/231601/introduction

[46] *Data Science Bowl 2017*. Accessed: Aug. 15, 2020. [Online]. Available: https://kaggle.com/c/data-science-bowl-2017

[47] *CodaLab—Competition*. Accessed: Aug. 15, 2020. [Online]. Available: http://isbichallenges.cloudapp.net/competitions/15#learn_the_details-overview

[48] *LNDb Challenge—Grand Challenge*. Accessed: Aug. 15, 2020. [Online]. Available: https://lndb.grand-challenge.org/TestRankingTeams/

[49] H. Macmahon, D. P. Naidich, J. M. Goo, K. S. Lee, A. N. C. Leung, J. R. Mayo, A. C. Mehta, Y. Ohno, C. A. Powell, M. Prokop, G. D. Rubin, C. M. Schaefer-Prokop, W. D. Travis, P. E. Van Schil, and A. A. Bankier, "Guidelines for management of incidental pulmonary nodules detected on CT images: From the fleischner society 2017," *Radiology*, vol. 284, no. 1, pp. 228–243, Jul. 2017, doi: 10.1148/radiol.2017161659.

[50] B. van Ginneken, "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study," *Med. Image Anal.*, vol. 14, no. 6, pp. 707–722, Dec. 2010, doi: 10.1016/j.media.2010.05.005.

[51] O. Ozdemir, R. L. Russell, and A. A. Berlin, "A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1419–1429, May 2020, doi: 10.1109/TMI.2019.2947595.

[52] R. Manickavasagam and S. Selvan, "Automatic detection and classification of lung nodules in CT image using optimized neuro fuzzy classifier with cuckoo search algorithm," *J. Med. Syst.*, vol. 43, no. 3, p. 77, Feb. 2019, doi: 10.1007/s10916-019-1177-9.

[53] H. Cheng, Y. Zhu, and H. Pan, "Modified U-Net block network for lung nodule detection," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, May 2019, pp. 599–605, doi: 10.1109/ITAIC.2019.8785445.

[54] A. Masood, "Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–13, 2020, doi: 10.1109/JTEHM.2019.2955458.

[55] F. Li, H. Huang, Y. Wu, C. Cai, Y. Huang, and X. Ding, "Lung nodule detection with a 3D ConvNet via IoU self-normalization and maxout unit," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1214–1218, doi: 10.1109/ICASSP.2019.8683537.

[56] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 991–1004, Apr. 2019, doi: 10.1109/TMI.2018.2876510.

[57] A. Kumar, M. Fulham, D. Feng, and J. Kim, "Co-learning feature fusion maps from PET-CT images of lung cancer," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 204–217, Jan. 2020, doi: 10.1109/TMI.2019.2923601.

[58] W. Zuo, F. Zhou, Z. Li, and L. Wang, "Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection," *IEEE Access*, vol. 7, pp. 32510–32521, 2019, doi: 10.1109/ACCESS.2019.2903587.

[59] O. Gunaydin, M. Gunay, and O. Sengel, "Comparison of lung cancer detection algorithms," in *Proc. Sci. Meeting Elect.-Electron. Biomed. Eng. Comput. Sci. (EBBT)*, Apr. 2019, pp. 1–4, doi: 10.1109/EBBT.2019.8741826.

[60] R. V. M. D. Nobrega, S. A. Peixoto, S. P. P. da Silva, and P. P. R. Filho, "Lung nodule classification via deep transfer learning in CT lung images," in *Proc. IEEE 31st Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2018, pp. 244–249, doi: 10.1109/CBMS.2018.00050.

[61] L. Liu, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Multi-task deep model with margin ranking loss for lung nodule analysis," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 718–728, Mar. 2020, doi: 10.1109/TMI.2019.2934577.

[62] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3484–3495, Nov. 2019, doi: 10.1109/TNNLS.2019.2892409.

[63] A. Soliman, F. Khalifa, A. Elnakib, M. Abou El-Ghar, N. Dunlap, B. Wang, G. Gimel'farb, R. Keynton, and A. El-Baz, "Accurate lungs segmentation on CT chest images by adaptive appearance-guided shape modeling," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 263–276, Jan. 2017, doi: 10.1109/TMI.2016.2606370.

[64] H. Shaziya, K. Shyamala, and R. Zaheer, "Automatic lung segmentation on thoracic CT scans using U-Net convolutional network," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2018, pp. 0643–0647, doi: 10.1109/ICCSP.2018.8524484.

[65] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. P. Naidich, and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Med.*, vol. 25, no. 6, pp. 954–961, Jun. 2019, doi: 10.1038/s41591-019-0447-x.

[66] J. Pedrosa, G. Aresta, C. Ferreira, M. Rodrigues, P. Leitão, A. S. Carvalho, J. Rebelo, E. Negrão, I. Ramos, A. Cunha, and A. Campilho, "LNDb: A lung nodule database on computed tomography," 2019, *arXiv:1911.08434*. [Online]. Available: http://arxiv.org/abs/1911.08434

[67] I. Bonavita, X. Rafael-Palou, M. Ceresa, G. Piella, V. Ribas, and M. A. G. Ballester, "Integration of convolutional neural networks for pulmonary nodule malignancy assessment in a lung cancer classification pipeline," *Comput. Methods Programs Biomed.*, vol. 185, Mar. 2020, Art. no. 105172, doi: 10.1016/j.cmpb.2019.105172.

[68] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 406–415, doi: 10.1109/ICCVW.2019.00052.

[69] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, Jan. 2019, Art. no. 014006, doi: 10.1117/1.JMI.6.1.014006.

[70] S. G. Armato, F. Li, M. L. Giger, H. Macmahon, S. Sone, and K. Doi, "Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program," *Radiology*, vol. 225, no. 3, pp. 685–692, Dec. 2002, doi: 10.1148/radiol.2253011376.

[71] S. A. Khan, S. Hussain, S. Yang, and K. Iqbal, "Effective and reliable framework for lung nodules detection from CT scan images," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 4989, doi: 10.1038/s41598-019-41510-9.

[72] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sanchez, and B. van Ginneken, "Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016, doi: 10.1109/TMI.2016.2536809.

[73] S. A. El-Regaily, M. A. M. Salem, M. H. Abdel Aziz, and M. I. Roushdy, "Multi-view convolutional neural network for lung nodule false positive reduction," *Expert Syst. Appl.*, Oct. 2019, Art. no. 113017, doi: 10.1016/j.eswa.2019.113017.

[74] S. V. Fotin, D. F. Yankelevitz, C. I. Henschke, and A. P. Reeves, "A multiscale Laplacian of Gaussian (LoG) filtering approach to pulmonary nodule detection from whole-lung CT scans," 2019, *arXiv:1907.08328*. [Online]. Available: http://arxiv.org/abs/1907.08328

[75] J. Wang, J. Wang, Y. Wen, H. Lu, T. Niu, J. Pan, and D. Qian, "Pulmonary nodule detection in volumetric chest CT scans using CNNs-based Nodule-Size-Adaptive detection and classification," *IEEE Access*, vol. 7, pp. 46033–46044, 2019, doi: 10.1109/ACCESS.2019.2908195.

[76] J. Liu, L. Cao, O. Akin, and Y. Tian, "Accurate and robust pulmonary nodule detection by 3D feature pyramid network with self-supervised feature learning," 2019, *arXiv:1907.11704*. [Online]. Available: http://arxiv.org/abs/1907.11704

[77] B. Liu, W. Chi, X. Li, P. Li, W. Liang, H. Liu, W. Wang, and J. He, "Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: Three decades' development course and future prospect," *J. Cancer Res. Clin. Oncol.*, vol. 146, no. 1, pp. 153–185, Jan. 2020, doi: 10.1007/s00432-019-03098-5.

[78] H. Tang, X. Liu, and X. Xie, "An End-to-End framework for integrated pulmonary nodule detection and false positive reduction," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 859–862, doi: 10.1109/ISBI.2019.8759244.

[79] S. Zheng, L. J. Cornelissen, X. Cui, X. Jing, R. N. J. Veldhuis, M. Oudkerk, and P. M. A. van Ooijen, "Efficient convolutional neural networks for multi-planar lung nodule detection: Improvement on small nodule identification," 2020, *arXiv:2001.04537*. [Online]. Available: http://arxiv.org/abs/2001.04537

[80] S. Liu, A. Arindra Adiyoso Setio, F. C. Ghesu, E. Gibson, S. Grbic, B. Georgescu, and D. Comaniciu, "No surprises: Training robust lung nodule detection for low-dose CT scans by augmenting with adversarial attacks," 2020, *arXiv:2003.03824*. [Online]. Available: http://arxiv.org/abs/2003.03824

[81] S. Zheng, J. Guo, X. Cui, R. N. J. Veldhuis, M. Oudkerk, and P. M. A. van Ooijen, "Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 797–805, Mar. 2020, doi: 10.1109/TMI.2019.2935553.

[82] G. Cao, T. Huang, K. Hou, W. Cao, P. Liu, and J. Zhang, "3D convolutional neural networks fusion model for lung nodule detection onClinical CT scans," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 973–978, doi: 10.1109/BIBM.2018.8621468.

[83] C. Jacobs, E. M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J.-M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop, and B. van Ginneken, "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Med. Image Anal.*, vol. 18, no. 2, pp. 374–384, Feb. 2014, doi: 10.1016/j.media.2013.12.001.

[84] A. Teramoto and H. Fujita, "Automated Lung Nodule Detection Using Positron Emission Tomography/Computed Tomography," in *Artificial Intelligence in Decision Support Systems for Diagnosis in Medical Imaging*, K. Suzuki and Y. Chen, Eds. Cham, Switzerland: Springer, 2018, pp. 87–110.

[85] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1558–1567, Jul. 2017, doi: 10.1109/TBME.2016.2613502.

[86] B.-C. Kim, J. S. Yoon, J.-S. Choi, and H.-I. Suk, "Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection," *Neural Netw.*, vol. 115, pp. 1–10, Jul. 2019, doi: 10.1016/j.neunet.2019.03.003.

[87] Z. Xiao, N. Du, L. Geng, F. Zhang, J. Wu, and Y. Liu, "Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images," *Appl. Sci.*, vol. 9, no. 16, p. 3261, Aug. 2019, doi: 10.3390/app9163261.

[88] W. Zuo, F. Zhou, and Y. He, "An embedded multi-branch 3D convolution neural network for false positive reduction in lung nodule detection," *J. Digit. Imag.*, vol. 12, pp. 1–12, Feb. 2020, doi: 10.1007/s10278-020-00326-0.

[89] H. Shakir, H. Rasheed, and T. M. Rasool Khan, "Radiomic feature selection for lung cancer classifiers," *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 5847–5855, May 2020, doi: 10.3233/JIFS-179672.

[90] N. Nasrullah, J. Sang, M. S. Alam, M. Mateen, B. Cai, and H. Hu, "Automated lung nodule detection and classification using deep learning combined with multiple strategies," *Sensors*, vol. 19, no. 17, p. 3722, Aug. 2019, doi: 10.3390/s19173722.

[91] Y. Balagurunathan, M. B. Schabath, H. Wang, Y. Liu, and R. J. Gillies, "Quantitative imaging features improve discrimination of malignancy in pulmonary nodules," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 8528, doi: 10.1038/s41598-019-44562-z.

[92] M. Al-Shabi, B. L. Lan, W. Y. Chan, K.-H. Ng, and M. Tan, "Lung nodule classification using deep Local–Global networks," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 14, no. 10, pp. 1815–1819, Oct. 2019, doi: 10.1007/s11548-019-01981-7.

[93] I. D. Apostolopoulos, "Experimenting with convolutional neural network architectures for the automatic characterization of solitary pulmonary Nodules' malignancy rating," 2020, *arXiv:2003.06801*. [Online]. Available: http://arxiv.org/abs/2003.06801

[94] J. Yang, H. Deng, X. Huang, B. Ni, and Y. Xu, "Relational learning between multiple pulmonary nodules via deep set attention transformers," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, p. 1875, doi: 10.1109/ISBI45749.2020.9098722.

[95] F. Ciompi, K. Chung, S. J. van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. W. Wille, A. Marchianá, U. Pastorino, M. Prokop, and B. van Ginneken, "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," *Sci. Rep.*, vol. 7, no. 1, Apr. 2017, Art. no. 46479, doi: 10.1038/srep46479.

[96] R. Gao, "Distanced LSTM: Time-distanced gates in long short-term memory models for lung cancer detection," in *Proc. 10th Int. Workshop*, 2019, pp. 310–318, doi: 10.1007/978-3-030-32692-0_36.

[97] J. L. Causey, J. Zhang, S. Ma, B. Jiang, J. A. Qualls, D. G. Politte, F. Prior, S. Zhang, and X. Huang, "Highly accurate model for prediction of lung nodule malignancy with CT scans," *Sci. Rep.*, vol. 8, no. 1, Jun. 2018, Art. no. 9286, doi: 10.1038/s41598-018-27569-w.

[98] S. Chen, K. Ma, and Y. Zheng, "Med3D: Transfer learning for 3D medical image analysis," 2019, *arXiv:1904.00625*. [Online]. Available: http://arxiv.org/abs/1904.00625

[99] D. Wang, Y. Zhang, K. Zhang, and L. Wang, "FocalMix: Semi-supervised learning for 3D medical image detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3950–3959, doi: 10.1109/CVPR42600.2020.00401.

[100] B. K. J. Veronica, "An effective neural network model for lung nodule detection in CT images with optimal fuzzy model," *Multimedia Tools Appl.*, vol. 79, nos. 19–20, pp. 14291–14311, May 2020, doi: 10.1007/s11042-020-08618-x.

[101] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[102] S. G. Armato, III, L. M. Hadjiiski, G. D. Tourassi, K. Drukker, M. L. Giger, F. Li, G. Redmond, K. Farahani, J. S. Kirby, and L. P. Clarke, "Special Section Guest Editorial: LUNGx Challenge for computerized lung nodule classification: Reflections and lessons learned," *JMI*, vol. 2, no. 2, Jun. 2015, Art. no. 020103, doi: 10.1117/1.JMI.2.2.020103.

[103] M. Dolejsi, J. Kybic, M. Polovincak, and S. Tuma, "The lung time: Annotated lung nodule dataset and nodule detection framework," *Proc. SPIE*, vol. 2013, Oct. 2013, Art. no. 72601U.

[104] R. H. Mak, M. G. Endres, and J. H. Paik, "Use of crowd innovation to develop an artificial intelligence-based solution for radiation therapy targeting," *JAMA Oncol.*, vol. 5, pp. 654–661, Apr. 2019.

[105] Z. Zhou, "Models genesis: Generic autodidactic models for 3D medical image analysis," in *Medical Image Computing and Computer Assisted Intervention*, vol. 11767, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P. T. Yap, and A. Khan, Eds. Cham, Switzerland: Springer, 2019, pp. 384–393.

[106] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and pancreatic tumor characterization in the deep learning era: Novel supervised and unsupervised learning approaches," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1777–1787, Aug. 2019, doi: 10.1109/TMI.2019.2894349.

[107] N. O'Mahony, "Deep learning vs. Traditional computer vision," in *Advances in Computer Vision*, vol. 943, K. Arai and S. Kapoor, Eds. Cham, Switzerland: Springer, 2020, pp. 128–144.

[108] Z. Shi, H. Hao, M. Zhao, Y. Feng, L. He, Y. Wang, and K. Suzuki, "A deep CNN based transfer learning method for false positive reduction," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 1017–1033, Jan. 2019, doi: 10.1007/s11042-018-6082-6.

[109] S. Ziyad, V. Radha, and V. Thavavel, "Performance evaluation of contrast enhancement techniques in computed tomography of lung images," in *Proc. IEEE 5th Int. Conf. for Converg. Technol.*, Mar. 2019, pp. 1–4, doi: 10.1109/I2CT45611.2019.9033602.

[110] C. Han, Y. Kitamura, A. Kudo, A. Ichinose, L. Rundo, Y. Furukawa, K. Umemoto, Y. Li, and H. Nakayama, "Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection," 2019, *arXiv:1906.04962*. [Online]. Available: http://arxiv.org/abs/1906.04962

[111] M. Kawagishi, T. Kubo, R. Sakamoto, M. Yakami, K. Fujimoto, G. Aoyama, Y. Emoto, H. Sekiguchi, K. Sakai, Y. Iizuka, M. Nishio, H. Yamamoto, and K. Togashi, "Automatic inference model construction for computer-aided diagnosis of lung nodule: Explanation adequacy, inference accuracy, and experts' knowledge," *PLoS ONE*, vol. 13, no. 11, Nov. 2018, Art. no. e0207661, doi: 10.1371/journal.pone.0207661.

WENMING CAO (Senior Member, IEEE) received the M.S. degree from the System Science Institute, Chinese Academy of Sciences, Beijing, China, in 1991, and the Ph.D. degree from the School of Automation, Southeast University, Nanjing, China, in 2003. From 2005 to 2007, he was a Postdoctoral Researcher with the Institute of Semiconductors, Chinese Academy of Sciences. He is currently a Professor with Shenzhen University, Shenzhen, China. His research interests include pattern recognition, image processing, and visual tracking.
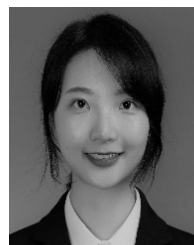
RUI WU received the B.Eng. degree in electronic information engineering from Shantou University, Shantou, China, in 2018. She is currently pursuing the master's degree in information and communication engineering with Shenzhen University, Shenzhen, China. Her current research interests include deep learning and medical image analysis.

GUITAO CAO (Member, IEEE) received the M.S. degree from Shandong University, Jinan, China, in 2001, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2006. She is currently an Associate Professor with the School of Computer Science and Software Engineering, East China Normal University, Shanghai. Her research interests include image processing and pattern recognition, media analysis, and understanding.

ZHIHAI HE (Fellow, IEEE) was a Research Engineer with the David Sarnoff Research Center. He is currently a Professor with the Electrical Engineering and Computer Science Department, University of Missouri. He was named as a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), in 2015, for his contributions to video communication and visual sensing technologies.

• • •