Computer-aided detection of lung nodules: a review

Furqan Shaukat
Gulistan Raja
Alejandro F. Frangi
Computer-aided detection of lung nodules: a review

Furqan Shaukat, a, * Gulistan Raja, a and Alejandro F. Frangi b

aUniversity of Engineering and Technology, Department of Electrical Engineering, Taxila, Pakistan
bUniversity of Leeds Woodhouse Lane, School of Computing and School of Medicine, Leeds, United Kingdom

Abstract. We present an in-depth review and analysis of salient methods for computer-aided detection of lung nodules. We evaluate the current methods for detecting lung nodules using literature searches with selection criteria based on validation dataset types, nodule sizes, numbers of cases, types of nodules, extracted features in traditional feature-based classifiers, sensitivity, and false positives (FP)/scans. Our review shows that current detection systems are often optimized for particular datasets and can detect only one or two types of nodules. We conclude that, in addition to achieving high sensitivity and reduced FP/scans, strategies for detecting lung nodules must detect a variety of nodules with high precision to improve the performances of the radiologists. To the best of our knowledge, ours is the first review of the effectiveness of feature extraction using traditional feature-based classifiers. Moreover, we discuss deep-learning methods in detail and conclude that features must be appropriately selected to improve the overall accuracy of the system. We present an analysis of current schemes and highlight constraints and future research areas. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.6.2.020901]

Keywords: computer-aided detection; lung nodule detection; lung cancer; false positive.

1 Introduction

Lung cancer is currently one of the most common causes of death worldwide, with low rates of survival after diagnosis being reported in developed and underdeveloped countries. According to recent statistics, the 5-year survival rate is only 16% and it has been estimated that by the year 2020, 12 million cancer-related fatalities will occur annually, of which lung cancer will have the largest share. However, survival rates can be improved if nodules are detected early enough. Lung nodules are abnormal growths of tissue that could represent lung cancer. They are typically round/spherical in shape with diameters of up to 30 mm. Nodules are categorized as well circumscribed, juxtavascular, juxtpleural, and pleural tail. Well-circumscribed nodules are independent and have no extensions into the surrounding anatomical structures, whereas juxtavascular nodules exhibit strong adherence to proximal vessels, and juxtopleural nodules are attached to neighboring pleural surfaces. Pleural-tail nodules have tails that are adherent to the nodule but not to pleural walls. In addition, pulmonary nodules are categorized as solid and subsolid nodules (SSNs), irrespective of their positions. SSNs are further classified as part-solid nodules and pure ground-glass nodules. Solid nodules are the most common type of nodules, and these repress the underlying functional lung tissues. SSNs are pulmonary nodules with partial ground-glass opacity (GGO). These nodules exhibit opacifications with higher density than the surrounding tissues and do not obscure underlying bronchovascular structures. Sample images of different nodules are shown in Fig. 1.

Computer-aided detection (CAD) can assist early diagnosis of lung cancer. The principal aim of CAD is to identify and to accurately extract regions of interest (ROIs) in images acquired from various imaging modalities, including computed tomography (CT), position emission tomography (PET), and magnetic resonance imaging (MRI). CAD systems can be further categorized as (i) computer-aided detection (CAD) and (ii) computer-aided diagnosis (CAD). The scope of CAD systems is limited to identification of suspicious areas in images, whereas CAD systems facilitate disease diagnosis. In this paper, we focus on CAD systems. A complete schematic representation of lung CAD processes is shown in Fig. 1.

Medical images are acquired from various imaging modalities. Among these, CT is a fundamental imaging technique for screening analyses of lung nodules, and the other available methods are of less importance. Among public databases, such as the Early Lung Cancer Action Program (ELCAP), Public Lung Image Database, and Public Lung Database to Address Drug Response, the Lung Image Database Consortium (LIDC) is widely used for its images because they carry standard radiological annotations. The most commonly used public databases are summarized in Table 1.

Other platforms that have contributed datasets to the research community include the Dutch–Belgian randomized lung cancer screening trial NELSON and the Lung Cancer Alliance. The main objective of these publicly available databases is to provide data resources to the research community for the development, evaluation, and benchmarking of CAD systems.

Lung segmentation is a process by which lung volumes are extracted from CT images and insignificant constituents are discarded. The efficiency of lung nodule detection systems is increased by accurate lung segmentation, and several techniques for extracting lung volumes from CT images are used. These include, optimal thresholding, rule-based region growing, global thresholding, three-dimensional (3-D) adaptive fuzzy thresholding, hybrid segmentation, and connected component labeling. Following preliminary lung segmentation, juxtopleural nodules are added by refining extracted lung volumes, generally using a chair-code method, a rolling ball algorithm, or morphological approaches.
Nodule detection can be described as a process in which suspicious lung areas are detected that may be responsible for lung cancer. Among the reported techniques for detecting lung nodules as candidate lung cancers, multiple gray-level thresholding is the most widely considered, although shape-based, template-matching-based, morphological approaches with convexity models, and filtering-based methods have also been used for this purpose.

Following detection of candidate nodules, nodules must be distinguished from non-nodules. In published studies, this false-positive (FP) reduction involves feature extraction and nodule classification using feature-based classifiers. Various methods are reported for extracting image features and classifying nodules, generally based on intensity-based statistical features, geometric features, and gradient features. After feature extraction, nodule detection is performed using several supervised and unsupervised classifiers to reduce the numbers of FPs. However, developments in deep learning have made the selection of image features less explicit, and optimal loss functions and efficient optimization algorithms that influence the learning process have been favored.

In Sec. 2, we present a review of studies that have been selected for their relevance to CADe. We have only considered studies from 2009 because the approaches reported prior to this time have become redundant. Our analyses of these studies are presented with a focus on limitations. Abstracts have been retrieved from PubMed, Science Direct, IEEE Xplore, and Web of Science using the keywords “lung,” “nodule,” “detection,” “pulmonary,” “tumor,” “CAD,” “CADe,” and “cancer” with various combinations of logical expressions containing “AND” and “OR.” We have reviewed only peer-reviewed archival journal publications and have included key conference papers that have been published in the past year. Section 3 presents a discussion of the major constraints on present and future prospects. Conclusions are drawn in Sec. 4. The aim of this review is to provide a critical analysis of current lung nodule detection systems and to highlight the constraints and future research areas.

## 2 Review of Lung Nodule Detection Systems

Lung nodule detection systems comprise processes for (i) lung segmentation, (ii) nodule candidate detection, and (iii) FP reduction. Several reviews of the methods used for nodule detection
and FP reduction identify overall sensitivity and numbers of FPs/scans as key performance criteria, but few comparative analyses have been performed to determine the effectiveness of the extracted features that are used for FP reduction. Therefore, we summarize the techniques for extracting features using feature-based classifiers that are used to determine the most relevant feature classes in lung nodule detection systems and to facilitate sensitivity and reduce FPs/scans of the system. Furthermore, we review reports of deep-learning techniques and compare their outcomes with those of traditional feature-based techniques. To the best of our knowledge, our review is comprehensive and up to date and comprises developments in the field. The present review highlights the challenges and constrains of the three categories of lung nodule detection system.

2.1 Lung Segmentation

Lung segmentation techniques can be broadly classified as (i) deformable boundary-based techniques, (ii) edge-based techniques, and (iii) threshold-based techniques. Each lung segmentation technique has its own pros and cons. Although threshold-based techniques are efficacious with high-contrast CT images, their performance can vary with low-contrast pathologies. Moreover, thresholding can be affected by differing imaging protocols and image acquisition scanners. Particularly, because lung structures, such as blood vessels, bronchioles, and bronchi, exhibit close densities with chest tissues, it is extremely challenging to accurately define ROIs and accurate segmentation often requires special postprocessing techniques. Deformable boundary-based techniques have the disadvantage of extra sensitivity to initialization. Furthermore, they are unable to overcome the heterogeneity of lung volumes with traditional external forces, such as edges and gray levels. Therefore, accurate lung segmentation is difficult using the deformable model. In addition, the accuracy of these imaging analyses depends on the accuracy of registration of prior shape models for CT images. Poor registration can affect the overall performance and is the main limitation of these schemes. In addition, the diversity of lung pathologies complicates the accurate segmentation of lung fields. Selected reports of lung segmentation techniques are summarized in Table 1.

2.2 Nodule Candidate Detection

Nodule candidate detection is performed to identify structures within the lung that are suspicious of being lung nodules. This process is typically performed following lung segmentation to decrease the workload by removing the background and unwanted areas from the input CT images. Various methods have been described for detecting lung nodule candidates, and multiple gray-level thresholding is considered the best method, although shape-based, template-matching-based, morphological approaches with convexity models, and filtering-based methods have been used.

Akram et al. applied multiple gray-level thresholding for nodule candidate detection and stated that single threshold values are insufficient because vessels and different types of nodules exhibit different density values. Choi and Cho reported that nodules exhibit a circular or dot-like shape of variable size. The authors suggested that single-scale enhancement is not appropriate for all nodules and reported the use of a multiscale dot enhancement filter. After enhancement, lung nodules were detected using thresholding. Gonçalves et al. and Chen et al. reported the use of Hessian matrix-based approaches for lung nodule detection. Gonçalves et al. used the central adaptive medianness principle for lung nodule identification and segmentation with shape indices and curvedness properties. They validated their method with 569 solid nodules of the LIDC-IDRI dataset and demonstrated superior results compared with those obtained via manual segmentation by expert radiologists. Choi and Cho proposed an entropy-based lung nodule detection system involving three stages. In the first stage, CT images are divided into informative and noninformative blocks and the latter are filtered out. In the next step, candidate nodules are detected using informative blocks after enhancement using 3-D coherence-enhancing diffusion. Candidate nodules are then detected from enhanced informative image blocks using optimal thresholding. Finally, certain features are extracted from lung nodule candidates, and FP reduction is performed using a support vector machine (SVM).

In this section, studies have been grouped according to the template-matching methods for lung nodule candidate detection. Jo et al. proposed lung nodule detection systems using template matching and reported a method based on global rib matching and nodule template matching. In their global rib-matching analyses, the lungs were aligned at their centers and rigid registration was performed using coronal and sagittal maximum intensity projection images. In the second step, lung nodule candidates were detected using template matching based on density similarities and geometrical correlations between nodules and other neighboring structures. Moreover, El-Baz et al. used two-dimensional (2-D) and 3-D deformable templates and a genetic optimization algorithm to detect lung nodule candidates.

Various morphological approaches have been used to detect lung nodule candidates. Cascio et al. proposed a lung nodule detection method using 3-D mass spring model. In their system, region-growing and morphological operations for lung volume segmentation were used, and lung nodule candidates were detected using a 3-D mass spring model. The range of gray values and corresponding shape information from the model helped in identifying lung nodule candidates with greater accuracy. The authors validated their system using 84 scans obtained from the LIDC dataset. Soltaninejad et al. proposed a lung nodule detection scheme using active contours and a K-nearest neighbor (KNN) classifier. After performing lung volume segmentation using adaptive thresholding and morphological operations, the lung nodule candidates were detected using 2-D stochastic features, followed by extraction using active contour modeling. Finally, FPs were reduced using the KNN classifier. Jianhao et al. proposed a shape-based lung nodule detection method comprising the three main steps: modeling, break, and repair. Initially, ROIs were extracted and represented as a shape model using the marching cubes algorithm, and the problematic regions were identified and removed using principal curvature analyses, which can lead to inaccurate segmentation of objects. Finally, incomplete regions were fitted using interpolation and extrapolation with a radial basis function for smoothly estimating and repairing suspicious areas. Kubota et al. proposed a lung nodule detection method using morphological operations and convexity models. Initially, lung volumes were extracted using voxel transformation and figure ground separation. Subsequently, a Euclidean distance map was used to locate the seed point and then region growing was applied to identify candidate nodule regions. Finally, the authors segmented candidate
lung nodules using convex hull. Reported techniques for lung nodule detection are summarized in Table 2.

In summary, the most commonly used lung nodule detection techniques can be broadly classified into the three main categories: (i) thresholding, (ii) template matching, and (iii) morphological approaches. Thresholding-based results depend on the qualities of techniques for threshold adjustment. Template-matching techniques suffer from irregular shapes and diversities of lung nodule types. Template-matching methods also generally assume that nodules are spherical or cylindrical and are, hence, challenged by nodules that are attached to the pleura and vessels. Alternatively, morphological approaches suffer from low detection efficiency for lung wall nodules.

### 2.3 False-Positive Reduction

After detecting nodule candidates, they are classified into nodules and non-nodules. This step is commonly referred to as FP reduction and is performed using the following two broad categories of methods: (i) conventional feature-based classifiers and (ii) convolutional neural networks (CNNs). Conventional feature-based classification is performed using (i) feature extraction and (ii) nodule candidate classification techniques. Several methods for feature extraction and nodule candidate classification have been proposed. Below, we briefly review published studies in both of these categories and highlight the challenges inherent in the respective CAD systems.

#### Table 2: Review of lung segmentation techniques.

<table>
<thead>
<tr>
<th>CAD systems</th>
<th>Year</th>
<th>No. of cases</th>
<th>Image size</th>
<th>Proposed technique</th>
<th>Ground truth</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soliman et al</td>
<td>2017</td>
<td>105</td>
<td>512 x 512 x 270–450</td>
<td>Shape-based</td>
<td>75 manually traced scans</td>
<td>OM = 0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSC = 98.4%</td>
</tr>
<tr>
<td>Filho et al</td>
<td>2017</td>
<td>40 CT scans</td>
<td>512 * 512</td>
<td>Shape-based</td>
<td>Semiautomatic (manual + commercial software)</td>
<td>FM = 99.14%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>deformable model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shi et al</td>
<td>2016</td>
<td>23 CT scans</td>
<td>512 * 512</td>
<td>Thresholding</td>
<td>23 manually traced data</td>
<td>OM = 0.98</td>
</tr>
<tr>
<td>Dai et al</td>
<td>2015</td>
<td>NA</td>
<td>512 * 512' 368</td>
<td>Shape-based</td>
<td>Manually traced data</td>
<td>DSC = 0.98</td>
</tr>
<tr>
<td>Mansoor et al</td>
<td>2014</td>
<td>400 CT images</td>
<td>NA</td>
<td>Shape-based</td>
<td>400 manually traced data</td>
<td>OM = 0.95</td>
</tr>
<tr>
<td>Sun et al</td>
<td>2012</td>
<td>30 scans</td>
<td>512 x 512 x 424–642, 0.6–0.7 mm thin</td>
<td>Shape-based</td>
<td>30 manually corrected traced data</td>
<td>DSC = 0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AD = 0.84 mm</td>
</tr>
<tr>
<td>Sofka et al</td>
<td>2011</td>
<td>260 scans</td>
<td>0.5 to 5.0 mm</td>
<td>Shape-based</td>
<td>68 manually traced data</td>
<td>SCD = 1.95</td>
</tr>
<tr>
<td>Besbes and Paragios</td>
<td>2011</td>
<td>247 image radiographs</td>
<td>256 x 256, 1-mm thin</td>
<td>Shape-based</td>
<td>123 manually traced data</td>
<td>OM = 0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AD = 1.39 pixel</td>
</tr>
<tr>
<td>Annangi et al</td>
<td>2010</td>
<td>1130 image radiographs</td>
<td>128 x 128 and 256 x 256</td>
<td>Shape-based deformable model</td>
<td>1130 manually traced images</td>
<td>DSC = 0.88</td>
</tr>
<tr>
<td>El-Baz et al</td>
<td>2008</td>
<td>10 image datasets</td>
<td>512 x 512 x 182, 2.5 mm thin</td>
<td>Statistical MGRF model</td>
<td>1820 manually traced images</td>
<td>Accu. = 0.96</td>
</tr>
<tr>
<td>Shi et al</td>
<td>2008</td>
<td>247 image radiographs</td>
<td>256 x 256</td>
<td>Shape-based deformable model</td>
<td>247 manually traced images</td>
<td>OM = 0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AD = 1.78 pixel</td>
</tr>
<tr>
<td>Gao et al</td>
<td>2007</td>
<td>8 subjects</td>
<td>512 x 512 x 240</td>
<td>Thresholding</td>
<td>8 manually traced datasets</td>
<td>DSC = 0.99</td>
</tr>
<tr>
<td>Korfiatis et al</td>
<td>2007</td>
<td>23 scans</td>
<td>512 x 512</td>
<td>Wavelet edge detector</td>
<td>22 manually traced data</td>
<td>OM = 0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AD = 0.77 mm</td>
</tr>
<tr>
<td>Campadelli et al</td>
<td>2006</td>
<td>487 image radiographs</td>
<td>256 x 256</td>
<td>Spatial edge detector</td>
<td>487 manually traced data</td>
<td>Sen. = 0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spec. = 0.96</td>
</tr>
<tr>
<td>Sluimer et al</td>
<td>2005</td>
<td>26 scans</td>
<td>512 x 512, 0.75–2.0 mm</td>
<td>Shape-based</td>
<td>10 manually traced Data</td>
<td>OM = 0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AD = 1.48 mm</td>
</tr>
</tbody>
</table>

Note: NA, not available; OM, overlap measure is defined as the volume of the intersection divided by the volume of the union of two samples; DSC, dice similarity coefficient is used to compare the similarity of two samples; FM, F-measure denotes the harmonic mean of predictive value and sensitivity; RmsD, root-mean-square difference of the distance between the segmentation and the ground truth; SCD, symmetrical point-to-mesh comparison error; AD, mean absolute surface distance is defined as symmetric border positioning measure integrated along entire surfaces.
Cuenca et al. proposed a CAD system using an iris filter to detect isolated pulmonary nodules from CT images. The system achieved a sensitivity of 80% with 7.7 FPs/scans. The system could only detect one type of nodule, although the used dataset was quite small and contained only 77 nodules. In contrast, Murphy et al. used a large private dataset of 813 scans for the evaluation of their proposed system and achieved a sensitivity of 80% with 4.2 FPs/scans. They used local image features and the k-nearest-neighbor classification. Despite their large dataset, the sensitivity of their system was lower than that of the other reported systems. Similarly, Guo et al., Liu et al., Retico et al., and Messay et al. used small datasets comprising 29 scans (34 true nodules), 32 scans (33 solitary nodules), 42 scans (102 pleural nodules), and 84 scans (150 nodules), respectively, for the evaluation of their proposed systems. It is presumed that the performances of these systems will be poor in realistic scenarios with a broader range of nodule types seen in clinical scans.

Ozekes et al. proposed a computerized lung nodule detection method using 3-D feature extraction and learning-based algorithms. They claimed a sensitivity of up to 100%, but a FP rate of 44 per scan rendered the method inefficient. An automatic CAD system was proposed by Sousa et al. which used an optimized subset of eight features from a total of 24 initially extracted features. The system achieved a FP rate of 0.42 and a sensitivity of 84.84%. However, it was tested with only 33 nodules, making its performance susceptible to differing scenarios. Mabrouk et al. proposed a technique for automatically detecting lung nodules from CT images using two classifiers. A total of 22 image features were extracted for their model and feature selection was driven by Fisher’s scores. Although the system exhibited good performance with respect to detection of large nodules, it was not able to detect smaller nodules.

Assefa et al. proposed a nodule detection scheme based on template matching and multisresolution-based false reduction. Seven statistical and two intensity-based features were extracted for the FP reduction stage and the system performed at an 81% classification rate. However, this system also had a very high FP rate (35.15%), leading to disadvantages in terms of inefficiency. Choi et al. proposed a method based on hierarchical block classification in which sub-blocks of the image were constructed and entropy-based analysis was then used to select those with high entropy. The proposed system achieved a sensitivity of 95.28% with only 2.27 FPs/scans. This system had good overall performance but failed to detect all types of nodules. Tariq et al. proposed a computerized system for lung nodule detection from CT scan images using a neurofuzzy classifier; however, no standard datasets or performance metrics were used to evaluate its performance. Orozco et al. extracted eight texture features from histograms and a gray-level co-occurrence matrix, which were given as input to SVM for FP reduction. The system achieved a reliability index of 84% but was evaluated using a private dataset of only 38 scans with nodules. Tariq et al. detected pulmonary nodules using hybrid features: a total of 30 intensity-based and geometrical (2-D and 3-D) features were extracted and given as input to four different classifiers. Their system achieved a sensitivity of 89.6% but was evaluated using a private dataset comprising only 95 pulmonary nodules.

Teramoto et al. proposed a hybrid method for detecting pulmonary nodules using PET/CT. They used 100 PET/CT images to evaluate their method, which achieved a sensitivity of 83.0% with five FPs/scans. Although their system relied on a novel combination of CT/PET images, it did not achieve high sensitivity. Choi et al. introduced a 3-D shape-based feature descriptor to detect pulmonary nodules in CT images. The system was evaluated using the LIDC dataset with 148 nodules and achieved a sensitivity of 97.5% with 6.76 FPs/scans. Although it showed good performance overall, the FP/scan was unfavorable. Akram et al. reported a SVM-based classification of lung nodules using hybrid features from CT images. Similar to other studies, their system was validated with insufficient nodules to achieve the same performance under various scenarios. Other selected studies that used conventional feature-based classification are summarized in Table 3.

This section presents selected studies that used CNN for pulmonary nodule detection. Setio et al. proposed a multiview convolutional network-based lung nodule detection system with three dedicated detectors for large, subsolid, and solid nodules. The final detection step was performed using multiple streams of 2-D convolutional networks and a dedicated fusion method. This system was evaluated using 888 scans from the LIDC-IDRI dataset and achieved a detection sensitivity of 90.1% with only four FPs/scans. Anirudh et al. used a 3-D CNN to learn discriminative features for nodule detection. The proposed system was evaluated using 67 scans from the SPIE-LUNGx dataset and achieved a relatively lower sensitivity of 80% with 10 FPs/scans. Ding et al. proposed a lung nodule detection system based on deep CNNs. Their system involved the application of a region-based CNN for nodule detection on image slices and employed a 3-D CNN to reduce FPs. It was evaluated using the Lung Nodule Analysis Challenge (LUNA16) dataset and achieved a high sensitivity (94.4%) with only four

<table>
<thead>
<tr>
<th>CAD systems</th>
<th>Year</th>
<th>Detection technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akram et al.</td>
<td>2016</td>
<td>Multiple gray-level thresholding</td>
</tr>
<tr>
<td>Choi and Cho</td>
<td>2014</td>
<td>Multiscale dot enhancement filter</td>
</tr>
<tr>
<td>Chen et al.</td>
<td>2013</td>
<td>Entropy analysis</td>
</tr>
<tr>
<td>Jo et al.</td>
<td>2014</td>
<td>Template matching</td>
</tr>
<tr>
<td>El-Baz et al.</td>
<td>2013</td>
<td>Template matching and genetic algorithm</td>
</tr>
<tr>
<td>Cascio et al.</td>
<td>2012</td>
<td>Stable 3-D mass spring models</td>
</tr>
<tr>
<td>Soltaninejad et al.</td>
<td>2012</td>
<td>Active contour and K-NN classifier</td>
</tr>
<tr>
<td>Jiantao et al.</td>
<td>2011</td>
<td>Thresholding and geometric modeling</td>
</tr>
<tr>
<td>Kubota et al.</td>
<td>2011</td>
<td>Convexity model and morphological approach</td>
</tr>
<tr>
<td>Riccardi et al.</td>
<td>2011</td>
<td>3-D fast radial transform</td>
</tr>
<tr>
<td>Namin et al. and Murphy et al.</td>
<td>2010, 2007</td>
<td>Shape index</td>
</tr>
<tr>
<td>Ozekes et al.</td>
<td>2008</td>
<td>3-D template matching</td>
</tr>
</tbody>
</table>
Table 4 Performance comparison of different CAD systems.

<table>
<thead>
<tr>
<th>CAD systems</th>
<th>Data set</th>
<th>No. of cases</th>
<th>No. of nodules</th>
<th>Nodule size (mm)</th>
<th>Extracted features</th>
<th>Sensitivity (%)</th>
<th>FPR</th>
<th>Type of nodules</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guo et al.</td>
<td>Private</td>
<td>29</td>
<td>34</td>
<td>N/A</td>
<td>Shape</td>
<td>94.77</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sousa et al.</td>
<td>Private</td>
<td>N/A</td>
<td>33</td>
<td>3 to 40</td>
<td>Shape, texture, gradient, histogram, and spatial</td>
<td>84.84</td>
<td>0.42</td>
<td>Isolated, juxtapleural and juxtavascular</td>
<td>Used dataset is too small containing a low number of nodules.</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>Private</td>
<td>32</td>
<td>33</td>
<td>3 to 17</td>
<td>N/A</td>
<td>93.75</td>
<td>4.60</td>
<td>Juxtapleural</td>
<td></td>
</tr>
<tr>
<td>Orozco et al.</td>
<td>LIDC, ELCAP</td>
<td>128</td>
<td>75</td>
<td>2 to 30</td>
<td>Texture</td>
<td>84.00</td>
<td>7.00</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Tartar et al.</td>
<td>Private</td>
<td>63</td>
<td>95</td>
<td>2 to 20</td>
<td>Shape</td>
<td>89.60</td>
<td>7.90</td>
<td>Well-circumscribed, vascularized, juxtapleural, and pleural-tail</td>
<td></td>
</tr>
<tr>
<td>Messay et al.</td>
<td>LIDC</td>
<td>84</td>
<td>143</td>
<td>3 to 30</td>
<td>Shape, intensity, and gradient</td>
<td>82.66</td>
<td>3.00</td>
<td>Juxtavascular and juxtapleural</td>
<td>Systems underperform in terms of sensitivity/accuracy.</td>
</tr>
<tr>
<td>Murphy et al.</td>
<td>Private</td>
<td>813</td>
<td>1518</td>
<td>2 to 14</td>
<td>Shape index and curvedness</td>
<td>80.00</td>
<td>4.20</td>
<td>Nonsolid, part-solid, and solid</td>
<td></td>
</tr>
<tr>
<td>Retico et al.</td>
<td>Private</td>
<td>42</td>
<td>102</td>
<td>6 to 30</td>
<td>Morphological and texture</td>
<td>72.00</td>
<td>6.00</td>
<td>Pleural</td>
<td></td>
</tr>
<tr>
<td>Teramoto et al.</td>
<td>Private</td>
<td>100</td>
<td>103</td>
<td>4 to 30</td>
<td>Shape and intensity</td>
<td>83.00</td>
<td>5.00</td>
<td>Solitary</td>
<td></td>
</tr>
<tr>
<td>Gong et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>Intensity, shape, and texture</td>
<td>79.30</td>
<td>4.00</td>
<td>Solid and GGO</td>
<td></td>
</tr>
<tr>
<td>Bergholdt et al.</td>
<td>LIDC</td>
<td>243</td>
<td>690</td>
<td>3 to 30</td>
<td>Shape, intensity, and gradient</td>
<td>85.90</td>
<td>2.50</td>
<td>Juxtapleural</td>
<td></td>
</tr>
<tr>
<td>Opfer et al.</td>
<td>LIDC</td>
<td>91</td>
<td>N/A</td>
<td>3 to 30</td>
<td>Shape and intensity</td>
<td>78.00</td>
<td>2.00</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Sahiner et al.</td>
<td>Private, LIDC</td>
<td>85</td>
<td>241</td>
<td>3 to 19</td>
<td>Shape, statistical, and gradient</td>
<td>76.00</td>
<td>5.60</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Suzuki et al.</td>
<td>Private</td>
<td>63</td>
<td>121</td>
<td>4 to 27</td>
<td>N/A</td>
<td>80.30</td>
<td>4.80</td>
<td>Pure GGO, mixed GGO, and solid</td>
<td></td>
</tr>
<tr>
<td>Ozekes et al.</td>
<td>LIDC</td>
<td>11</td>
<td>11</td>
<td>3 to 16</td>
<td>Shape</td>
<td>100.00</td>
<td>44.00</td>
<td>Solitary</td>
<td>High FP rate makes the schemes inefficient.</td>
</tr>
<tr>
<td>Assefa et al.</td>
<td>ELCAP</td>
<td>50</td>
<td>165</td>
<td>N/A</td>
<td>Intensity and statistical</td>
<td>81.00</td>
<td>35.15</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Torres et al.</td>
<td>LIDC</td>
<td>949</td>
<td>1749</td>
<td>3 to 30</td>
<td>Shape and intensity</td>
<td>80.00</td>
<td>8.00</td>
<td>GGO</td>
<td></td>
</tr>
<tr>
<td>Choi et al.</td>
<td>LIDC</td>
<td>84</td>
<td>148</td>
<td>3 to 30</td>
<td>Shape-based 3-D descriptor</td>
<td>97.50</td>
<td>6.76</td>
<td>Solid and juxtapleural</td>
<td></td>
</tr>
<tr>
<td>Mabrouk et al.</td>
<td>Private</td>
<td>12</td>
<td>N/A</td>
<td>22 to 42</td>
<td>Shape and intensity</td>
<td>97.00</td>
<td>2.00</td>
<td>N/A</td>
<td>System's ability to detect all type of nodules is limited.</td>
</tr>
<tr>
<td>Choi et al.</td>
<td>LIDC</td>
<td>58</td>
<td>151</td>
<td>3 to 30</td>
<td>Shape and intensity</td>
<td>95.28</td>
<td>2.27</td>
<td>Juxtapleural</td>
<td></td>
</tr>
<tr>
<td>CAD systems</td>
<td>Data set</td>
<td>No. of cases</td>
<td>No. of nodules</td>
<td>Nodule size (mm)</td>
<td>Extracted features</td>
<td>Sensitivity (%)</td>
<td>FPR</td>
<td>Type of nodules</td>
<td>Remarks</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>--------------</td>
<td>----------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>------</td>
<td>-----------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Akram et al.</td>
<td>LIDC</td>
<td>47</td>
<td>50</td>
<td>3 to 30</td>
<td>Shape and intensity</td>
<td>95.31</td>
<td>N/A</td>
<td>Juxtapleural</td>
<td>System is evaluated with small number of nodules and FP/scan is not informed.</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>LIDC</td>
<td>1010</td>
<td>673</td>
<td>3 to 30</td>
<td>CS-LBP and ORT-EOH</td>
<td>95.69</td>
<td>3.05</td>
<td>Solid,GGO, juxtavascular, and juxtapleural</td>
<td></td>
</tr>
<tr>
<td>Setio et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>90.10</td>
<td>4.00</td>
<td>Solid, subsolid, and juxtapleural</td>
<td>Developments on deep-learning have made less explicit the selection of the image features, which has now turned to select optimal loss functions and efficient optimization algorithms influencing the learning process. In addition, CNN may have a high computational cost and requires a large dataset for training.</td>
</tr>
<tr>
<td>Anirudh et al.</td>
<td>SPIE-AAPM LUNG</td>
<td>67</td>
<td>N/A</td>
<td>3 to 30</td>
<td>CNN</td>
<td>80.00</td>
<td>10.00</td>
<td>Solid, part-solid and nonsolid</td>
<td></td>
</tr>
<tr>
<td>Ding et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>94.40</td>
<td>4.00</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Gruetzemacher et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>89.29</td>
<td>1.79</td>
<td>Juxtapleural and juxtavascular</td>
<td></td>
</tr>
<tr>
<td>Xie et al.</td>
<td>LIDC</td>
<td>1018</td>
<td>2669</td>
<td>3 to 30</td>
<td>CNN</td>
<td>84.19</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Kim et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1166</td>
<td>3 to 30</td>
<td>CNN</td>
<td>95.20</td>
<td>2.00</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Qin et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>98.2</td>
<td>4.00</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Xie et al.</td>
<td>LIDC</td>
<td>1018</td>
<td>N/A</td>
<td>3 to 30</td>
<td>CNN</td>
<td>83.2</td>
<td>4.00</td>
<td>Solitary, vascularized, juxtapleural, and pleural-tail</td>
<td></td>
</tr>
<tr>
<td>Dou et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>90.70</td>
<td>4.00</td>
<td>Solitary, GGO, and pleural</td>
<td></td>
</tr>
<tr>
<td>Jiang et al.</td>
<td>LIDC</td>
<td>1006</td>
<td>2669</td>
<td>3 to 30</td>
<td>CNN</td>
<td>80.06</td>
<td>4.70</td>
<td>Juxtapleural nodule</td>
<td></td>
</tr>
<tr>
<td>Jin et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>92.40</td>
<td>2.00</td>
<td>Solid, subsolid, and pleural</td>
<td></td>
</tr>
<tr>
<td>Dou et al.</td>
<td>LIDC</td>
<td>888</td>
<td>1186</td>
<td>3 to 30</td>
<td>CNN</td>
<td>90.60</td>
<td>2.00</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

Note: N/A, Not available.
FPs/scans. Zhu et al.99 developed the automatic lung nodule detection and classification system DeepLung, which included nodule detection and classification. Nodule detection was achieved using a 3-D fast regional CNN (R-CNN), and the system achieved a relatively lower detection sensitivity of 83.4%.

Gruetzemacher et al.96 proposed a lung nodule detection method using two 3-D CNNs: the first was used to generate candidate nodules and the second was used to reduce FPs. Using 888 scans from the LIDC dataset, a sensitivity of 89.29% was demonstrated with 1.78 FPs/scans. Xie et al.90 proposed a lung nodule detection method that employed different approaches for feature extraction. Feature representations of nodules were learned using deep CNN and candidate nodules were classified as nodules or non-nodules using the AdaBoost backpropagation neural network. The proposed system achieved a sensitivity of 84.19% with 7.98 FPs/scans. Similarly, Kim et al.92 proposed a lung nodule detection method using multiscale gradual integration of CNN in a three-step method. Multiscale patches with differing levels of contextual information were gradually integrated using zoom-in and zoom-out streams. The reported competition performance metric of 0.942 indicates an average sensitivity of analyses performed at seven different FP rates using the LIDC dataset for evaluation. In their CAD system, Qin et al.93 used a 3-D CNN model that employed 3-D U-Net architecture as the backbone for a region proposal network (RPN). It had a sensitivity of 98.2% with only four FPs/scans. Xie et al.94 contributed a 2-D CNN for pulmonary nodule detection. They detected nodule candidates by adjusting the structures of a faster R-CNN with two RPNs and a deconvolution layer. Their approach was extensively evaluated using the LIDC dataset used in the LUNA16 study and achieved a sensitivity of 86.4% with only four FPs/scans. The CAD systems described above are summarized in Table 4.

3 Discussion

To identify the challenges and future research directions, we summarized selected lung nodule detection systems reported in the literature since 2009. In this review of current methods, direct comparisons of research results were hampered by diverse performance metrics and evaluation protocols. Nonetheless, we evaluated the present systems according to the datasets used, number of subjects considered, nodule sizes, nodule numbers, and the standard performance metrics, sensitivity and FPs/scans. We also compared lung nodule features that were extracted in the reviewed studies, and identified the most relevant features for effective lung nodule detection systems. To this end, we grouped reported systems into the following categories:

i. Papers with small datasets and small numbers of nodules; the performance of these systems will likely deteriorate under more realistic scenarios with more various nodule types, as present in clinical scans.

ii. Papers reporting systems with poor accuracy/sensitivity compared with other systems.

iii. Papers in which high FP rates hamper efficiency.

References 44 and 46 were included in the table based on relevance, and the results of some other studies were omitted due to the absence of relevant information. Collectively, the studies included in Table 4 indicate that the major challenge for lung nodule detection systems is robustness to diverse clinical data of varying quality. In particular, most algorithms were optimized using private datasets, thus limiting comparability and generalization of the results. In addition, to ensure robustness, the proposed methods need to be validated with sufficiently large datasets that include all nodule types. Accordingly, methods that were evaluated with fewer nodules will likely lose accuracy under clinical conditions in which, nodule types are more varied. Feature extraction serves as an important step in differentiating nodules from other anatomic structures present in lung lobes. Yet, optimal set of features for nodule detection remain a subject of debate. The major constraints of lung nodule detection are summarized as follows:

i. Nodule detection methods are demonstrated using particular datasets.

ii. Few methods have been validated with large datasets.

iii. Optimal selection of features for nodule detection.

iv. Robustness to diverse nodule types.

v. Inconsistent use of performance metrics.

vi. Robustness to diverse lung nodule size.

Constraints that influence nodule detection remain a challenge in this area, in part because reported systems have been developed to accommodate the specific requirements of the investigating practitioners. The remaining challenge is to develop more accurate and robust systems that identify a broad range of nodules with increased sensitivity and reduced FPs/scans. Some of the present studies, however, have the potential to facilitate the development of lung cancer diagnosis tools. Specifically Choi et al.97 El-Baz et al.98 Mansoor et al.99 Dai et al.100 Soliman et al.101 Filho et al.102 Setio et al.103 Ding et al.104 and Shaukat et al.78 have achieved high performance metrics and have validated their methods using large public datasets, such as the LIDC. This database is distinguished by standard radiological annotations that have been generated by four expert radiologists in two consecutive sessions.

3.1 Future Prospects

Much further research is required to improve the CAD systems for lung cancer. Despite the considerable volume of research in this area, no commercial products are available for use in hospitals, reflecting the need for further research and development of the related technologies. The following critical topics can be identified in collective considerations of the present reviewed studies:

1. Segmentation of suspected pulmonary nodules requires further research and development. Accurate pulmonary nodule segmentation can increase the detection sensitivity of CAD systems.

2. CAD systems need to be validated with sufficiently large datasets to demonstrate robustness. Many CAD systems have only been evaluated on relatively small datasets, and their performance will likely be reduced in real clinical scenarios. More extensive experiments will provide assessments of the generalizability and clinical performance of these detection systems.
3. Selection of optimal features for lung nodule detection is another area needing further investigation. Although deep-learning technologies avoid handcrafting and selecting image features, they instead require selection of a loss function, network architecture, and an efficient optimization method, all of which influence the learning process.

4. Future CAD systems should be able to detect all types of nodules with the same precision and sensitivity and with reduced FPs/scans.

4 Conclusion
The existing methods for detecting lung nodules need to be improved, and this may be achieved by proposing new techniques and providing novel solutions. Future CAD$_2$ systems will be expected to detect all types of nodules with high precision and sensitivity and with few FPs/scans. To ensure robustness, proposed systems will need to be evaluated on large datasets so that evaluations of multiple datasets with standard performance metrics can be performed with accuracy. A well-performing CAD system would save many lives by facilitating early detection of lung nodules and providing a second opinion to that of expert radiologists.

Disclosures
The authors declare no conflicts of interest.

Acknowledgments
The authors would like to thank Enago (www.enago.com) for the English language review.

References
nODULES IN LOW-DOSE COMPUTED TOMOGRAPHY,” \textit{Med. Phys.} 30(7), 1602–
1617 (2003).
86. E. L. Torres et al., “Large scale validation of the M5L lung CAD on
87. A. A. Setio et al., “Pulmonary nodule detection in CT images: false
positive reduction using multi-view convolutional networks,” \textit{IEEE
88. R. Anirudh et al., “Large scale validation of the M5L lung CAD on
89. J. Ding et al., “Accurate pulmonary nodule detection in computed
tomography images using deep convolutional neural networks,”
90. R. Gruetzemacher, A. Gupta, and D. Paradice, “3D deep learning for
detecting pulmonary nodules in CT scans,” \textit{J. Am. Med. Inf. Assoc}
91. Y. Xie et al., “Fusing texture, shape and deep model-learned information
at decision level for automated classification of lung nodules on
92. B.-C. Kim et al., “Multi-scale gradual integration CNN for false
positive reduction in pulmonary node detection,” \textit{Neural Network}
115, 1–10 (2019).
93. Y. Qin et al., “Simultaneous accurate detection of pulmonary nodules
94. H. Xie et al., “Automated pulmonary nodule detection in CT images
using deep convolutional neural networks,” \textit{Pattern Recognit.}
95. Q. Dou et al., “Multilevel contextual 3-D CNNs for false positive
64(7), 1558–1567 (2017).
97. H. Jin et al., “A deep 3D residual CNN for false-positive reduction in
98. Q. Dou et al., “Automated pulmonary nodule detection via 3D
ConvNets with online sample filtering and hybrid-loss residual learn-
pulmonary nodule detection and classification,” arXiv:1709.05538
(2017).
100. S. A. Patil and V. R. Udupi, “Geometrical and texture features estima-
101. S. A. Kumar et al., “Robust and automated lung nodule diagnosis from
102. F. Zhang et al., “Overlapping node discovery for improving classifi-

Furqan Shaukat is currently working as an assistant professor in the
Department of Electronics Engineering at UET Taxila Sub Campus
Chakwal. He completed his BSc degree in electrical engineering at
UET Lahore in 2007. He received his MSc degree and his PhD
from UET Taxila in 2011 and 2018, respectively. He also worked
as a research associate at the University of Sheffield, U.K., during his
PhD. His research interests include medical image analysis and
classification.

Gulistan Raja received his BSc electrical engineering degree from
UET Taxila in 1996. He completed his master’s degree in information
systems engineering from Osaka University, Osaka, Japan, in 2002 and
his PhD in electrical engineering from UET Taxila in 2008. Currently, he is serving as a professor of electrical engineering at
UET Taxila. He has authored/coauthored more than 90 research pub-
lications in reputed international journals and refereed conferences.
His research interests include digital image/video signal processing
and VLSI design.

Alejandro F. Frangi is a Diamond Jubilee Chair in computational
medicine at the University of Leeds, Leeds, U.K., with joint appoint-
ments at the School of Computing and the School of Medicine. He
has been awarded a chair in emerging technologies by the Royal
Academy of Engineering. He leads the Center for Computational
Imaging and Simulation Technologies in Biomedicine. His main
research interests are in medical image computing, medical imaging,
and image-based computational physiology.