Research Article

A Computer-Aided Pipeline for Automatic Lung Cancer Classification on Computed Tomography Scans

Emre Dandıl [ID]

Department of Computer Engineering, Faculty of Engineering, Bilecik Seyh Edebali University, Gulumbe Campus, 11210 Bilecik, Turkey

Correspondence should be addressed to Emre Dandıl; emre.dandil@bilecik.edu.tr

Lung cancer is one of the most common cancer types. For the survival of the patient, early detection of lung cancer with the best treatment method is crucial. In this study, we propose a novel computer-aided pipeline on computed tomography (CT) scans for early diagnosis of lung cancer thanks to the classification of benign and malignant nodules. The proposed pipeline is composed of four stages. In preprocessing steps, CT images are enhanced, and lung volumes are extracted from the image with the help of a novel method called lung volume extraction method (LUVEM). The significance of the proposed pipeline is using LUVEM for extracting lung region. In nodule detection stage, candidate nodules are determined according to the circular Hough transform- (CHT-) based method. Then, lung nodules are segmented with self-organizing maps (SOM). In feature computation stage, intensity, shape, texture, energy, and combined features are used for feature extraction, and principal component analysis (PCA) is used for feature reduction step. In the final stage, probabilistic neural network (PNN) classifies benign and malign nodules. According to the experiments performed on our dataset, the proposed pipeline system can classify benign and malign nodules with 95.91% accuracy, 97.42% sensitivity, and 94.24% specificity. Even in cases of small-sized nodules (3–10 mm), the proposed system can determine the nodule type with 94.68% accuracy.

1. Introduction

Nowadays, lung cancer is one of the ranking first causes of mortality worldwide among men and women [1, 2]. Although there are a lot of treatment options like surgery, radiotherapy, and chemotherapy, five year survival rate for patients is quite low [3]. However, survival rate may go up to 54% in case lung cancer is identified in an early stage [4]. Therefore, early detection of lung cancer is vital to decrease lung cancer mortality.

Medical imaging techniques have been important technology in screening of lung cancer recently. CT scan becomes a standard modality for detecting and assessing lung cancer [5]. Most of the lung nodules are usually benign. However, some nodules such as calcified, swollen, and hard can also be determined as benign. Similarly, a hard nodule generally is cancerous (malignant), but it may be considered as benign case in some cases [6]. Furthermore, medical CT images are needed to be diagnosed by radiologists.

Computer-aided detection (CAD) systems have been an important field in medical image processing. CAD systems also based on machine learning methods designed to diagnosis of cancer have become common in recent years. Radiologists and physicians may use findings of CAD systems as the second opinion before making their own final decisions. Therefore, CAD systems play an important role in CT scans to help radiologists for detection of lung cancer efficiently.

2. Related Work

Computer-aided detection (CAD) systems have been active research field for the pulmonary nodule detection and malign/benign nodule classification. Until now, many CAD systems have been proposed. For example, Ozekes and Camurcu proposed a method for pulmonary nodule detection method using template matching [7]. Schilham et al. presented a CAD system which consists of image

preprocessing, candidate nodule detection, feature extraction, and classification for nodule detection in chest radiographs [8]. Dehmeshki et al. detected lung nodules using shape-based genetic algorithm template matching [9]. Suarez-Cuenca et al. also designed a system which discriminates the nodules and non-nodule cases using iris filter in CT images [10]. Murphy et al. automatically performed lung nodule detection using k-nearest neighbours classifier [11]. Giger et al. realized CAD system to detect lung nodules on CT images using geometric features [12]. In addition, Hasegawa et al. proposed image processing methods for identification of lung nodules using CT scans [13]. In another study, Kanazawa et al. used a CAD system to identify pulmonary nodules with fuzzy features [14]. In 2005, Suzuki et al. proposed a method using ANN for classification of malignant and benign nodules on CT images [15]. Sun et al. compared support vector machines (SVM) with the some classification methods for detection of lung cancer on CT images [16]. Kuruvilla and Gunavathi proposed a system using ANNs for classification of lung cancer [17].

In a recent study on lung nodule detection, Javaid et al. proposed a computer aided nodule detection method for the segmentation and detection of challenging in different type nodules [18]. ur Rehman et al. presented a systematic analysis of nodules detection techniques with the current trends and future challenges [19]. Wang et al. proposed a pulmonary nodule CAD based on semisupervised extreme learning machine [20]. Xie et al. proposed an automated pulmonary nodule detection system with 2D convolutional neural network (CNN) on LUNA16 dataset [21].

In this study, we have proposed fully automated computer-aided pipeline for the detection of pulmonary nodules and classification of benign/malign nodules in early stage. The contributions of this paper are (1) to review the systematic literature review; (2) to present the state of the art detection of pulmonary nodules and classification of lung cancer; (3) to propose the novel preprocessing method (LUVEM) for the lung volume extraction; (4) to suggest a novel candidate lung nodule detection method using CHT; (5) to design a holistic pipeline for the detection of pulmonary nodules as well as lung cancer; (6) the detailed comparison of feature extraction methods for lung nodule detection; and (7) to perform the detailed performance evaluation, high true detection rate, and low false detection rate for nodule detection and classification.

3. Architecture of the Computer-Aided Pipeline

Designed pipeline consists of four main stages such as image preprocessing (Stage I), lung nodule detection (Stage II), nodule feature computation (Stage III), and nodule classification (Stage IV). The work flow of the pipeline is presented in Figure 1.

3.1. Lung Image Preprocessing. In the first step of the image preprocessing stage, reading of CT images is performed. The CT scans obtained for the work are stored as DICOM

(Digital Imaging and Communications in Medicine) files [22].

The goal of image enhancement step is to prevent misleading results that may occur in subsequent processes. Thus, we firstly implemented the median filter to remove unnecessary noises and enhance the images. Moreover, the sharpening of nodule contours is an important step for the detection of nodules. Laplacian filter was used in our study. So, nodules on lung region were able to be detected more accurately. Furthermore, histogram equalization was also used in enhancement step in order to minimize contrast differences which occur due to scanning errors and to remove unnecessary grains.

In lung volume extraction step of image preprocessing stage, extracting of the lung region from CT image is performed. There are many methods for extracting lung volume from a lung CT scan [9, 10, 23–25]. However, these methods are complex, and they require more processing overhead. In some cases, these methods may lead to losses of information about lung regions or cause noise. The purpose of this step is to extract the lung region completely from the full lung CT image. Therefore, a simple but effective and novel method has been proposed in this study for lung volume extraction named as lung volume extraction method (LUVEM). The pseudocode of LUVEM is shown in Algorithm 1.

In LUVEM method, lung lobes are extracted from CT images with the help of morphological operations. LUVEM removes unrelated segments on the sides and edges of the preprocessed image and obtain the lung region successful. In the algorithm, input image is firstly converted to double-formatted image. Afterwards, 1 or 0 values are assigned to each pixel of double-formatted image according to low and high threshold values. The low and high threshold values are determined 0.25 and 0.65 in this algorithm, respectively. The method removes the bright areas on the edges of the lung CT image since their average values change between low and high values. After this process, the image is converted to binary format and performed morphological operations which are eroding, dilating, and filling, respectively. Finally, the image is again converted to gray-scale format. The segmentation examples of LUVEM can be seen in Figure 2. It is clearly seen that LUVEM can successfully extract the lung volume. In addition, quantitative evaluation of LUVEM will be reported below.

3.2. Lung Nodule Detection. The first step of lung nodule detection is candidate nodule detection. The nodule candidates in volume should be detected before nodule segmentation. The lung volume includes vessels and nodules. Moreover, the density of nodules, vessels, and lungs is different from each other [26]. Since the lung nodules have a circular and helical structure, they can be differentiated by means of circularity determination. Many methods have been suggested for identifying the round objects. Circular Hough transform (CHT), which proposed by Duda et al. [27], is one of the most successful method [28] for detection of round objects on the images. In this study, CHT
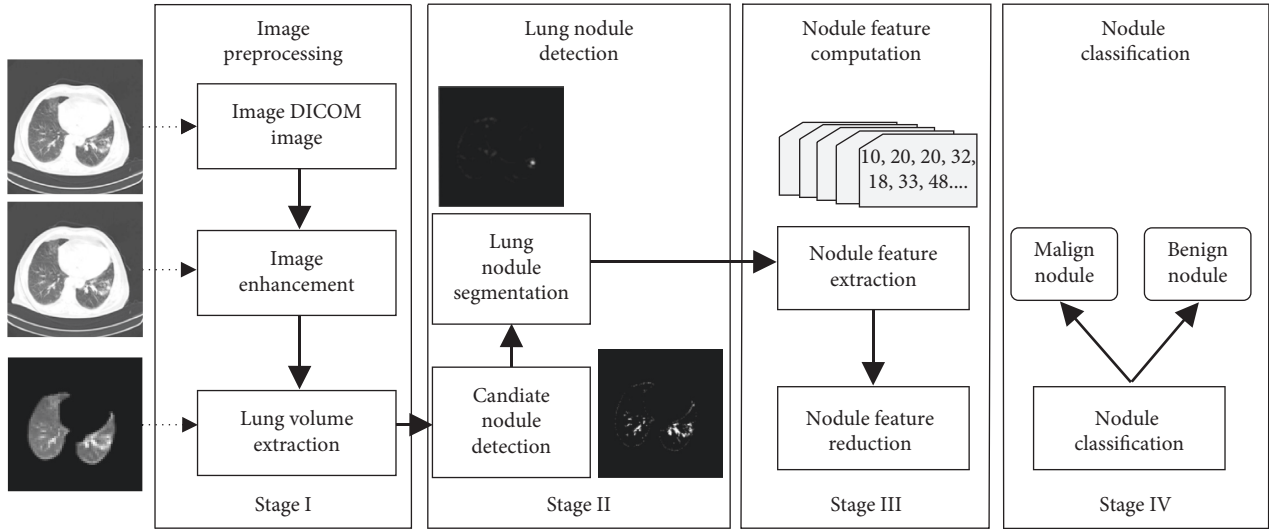
FIGURE 1: Work flow of the designed pipeline to detect lung cancer. The system consists of four stages: Stage I—enhancement of lung CT image and a novel lung volume segmentation method (LUVEM), Stage II—candidate nodule detection using CHT and segmentation of lung nodules using SOM, Stage III—computing of lung nodule features and reduction features using PCA, and Stage IV—classification of malign and benign lung nodule using PNN.

```
(1)    procedure LungVolumeExtraction(input image, low, high) //beginning lung vol. extraction algorithm
(2)        I ⟵ input image //input image
(3)        O ⟵ Ø, Lo ⟵ low, Hi ⟵ high //output image, low and high threshold
(4)            DIm ⟵ Converting (I, double) //converting image to double image
(5)        for (i ϵ row length of DIm) //starting loop of extraction step
(6)            for (j ϵ column length of DIm) //starting of inner loop
(7)                if DIm(i, j) ≥ Lo and DIm(i, j) ≤ Hi then //starting of if
(8)                    DIm(i, j) ⟵ 1; //assignment of binary 1
(9)                else
(10)                   DIm(i, j) ⟵ 0; //assignment of binary 0
(11)               end if //ending of if
(12)           end for //ending of inner loop
(13)       end for //ending loop of extraction step
(14)           BI ⟵ Converting(DIm, binary) //converting image to binary image
(15)           FI ⟵ Morphological eroding, dilating, filling (BI) //morphological operations
(16)           GIm ⟵ Converting(FI, gray-scale) //converting image to gray-scale image
(17)       O ⟵ GIm //return output image
(18)   end procedure //ending of LUVEM method
```

ALGORITHM 1: The pseudocode of lung volume extraction method (LUVEM).

operations are used for candidate nodule detection. CHT can detect the round object in the image; moreover, it can also detect the noncircular object by means of some operations. The image dataset is divided into 3 categories according to the nodule size such as <10 mm, 10–20 mm, and >20 mm. In order to detect the nodules in different size by CHT, three minimum and maximum radiuses such as 3–12 mm, 10–20 mm, and 15–45 mm are determined. In Figure 3, it is shown that the examples of determination of candidate nodules on CT images.

The second step of lung nodule detection is nodule segmentation. In this study, SOM [29] is proposed to segment nodules on CT images. SOM is an unsupervised neural network learning [30] method. It can perform on large/complex datasets [31, 32]. Furthermore, it designs data maps that can be interpreted easily. In addition to these advantages, SOM can easily segment very small nodules on the lung CT images [3]. The examples of segmented lung nodule images using SOM are shown in Figure 4.

3.3. Nodule Feature Computation.
Generally, CAD systems segment lung nodules for the determination of nodule candidates, and then features extract from the candidate nodules. The popular features are geometric feature, gray level features, gradient features, and energy level features. Therefore, we extracted 2D significative features from lung

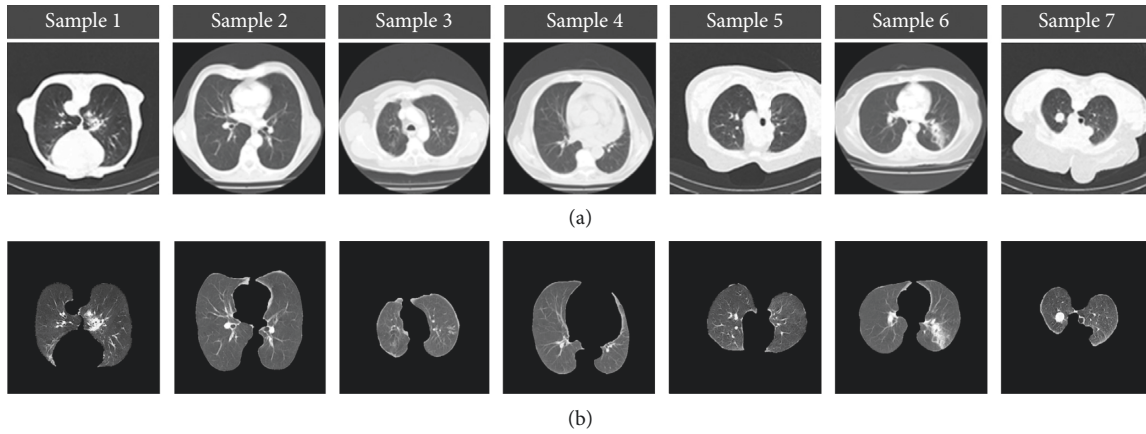| Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7

(a)

(b)

FIGURE 2: Extraction examples of lung regions: (a) preprocessed images and (b) lung volume extraction using LUVEM.
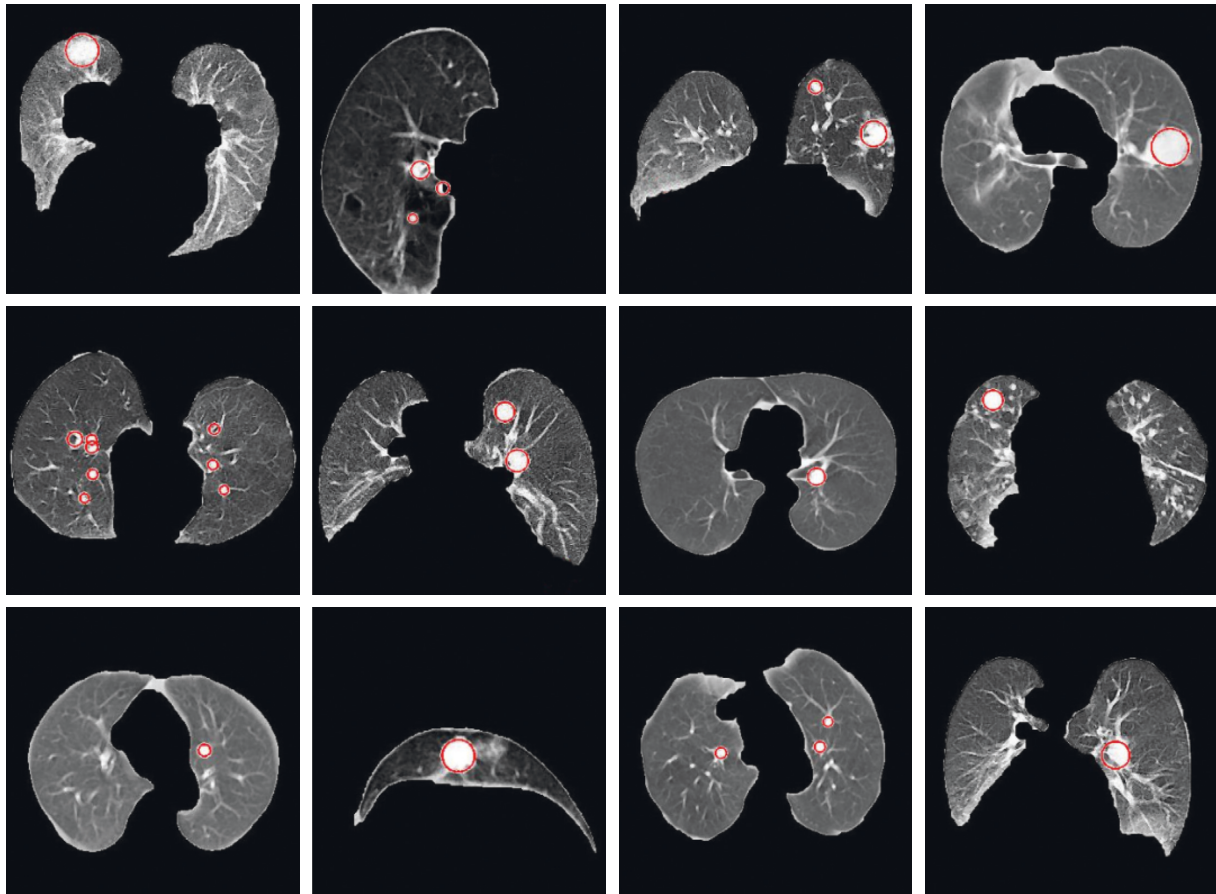
FIGURE 3: Determination of candidate lung nodule using CHT.

CT images to discriminate benign or malign nodule. Firstly, we used shape-based features for analyzing nodule geometry. We used first-order statistical features to obtain global statistic about nodule region. Moreover, we utilized gray level co-occurrence matrix (GLCM) texture features for gray level statistic of nodules. Finally, we extracted wavelet decomposition transform features to obtain the energy feature of nodules. All computed features are extracted from the slice of the segmented object.

First-order statistical features (SSF) of an image are calculated from the gray level histogram values of an image [33]. In this study, 6 basic features such as standard deviation, entropy, means, skewness, kurtosis, and variance were extracted by SSF using the histogram values of a gray level lung CT image. Shape-based features (SBF) allow feature extraction from the image by using geometric parameters [34]. Shape features give some information about an image such as sharpness, circularity, and convexity. In
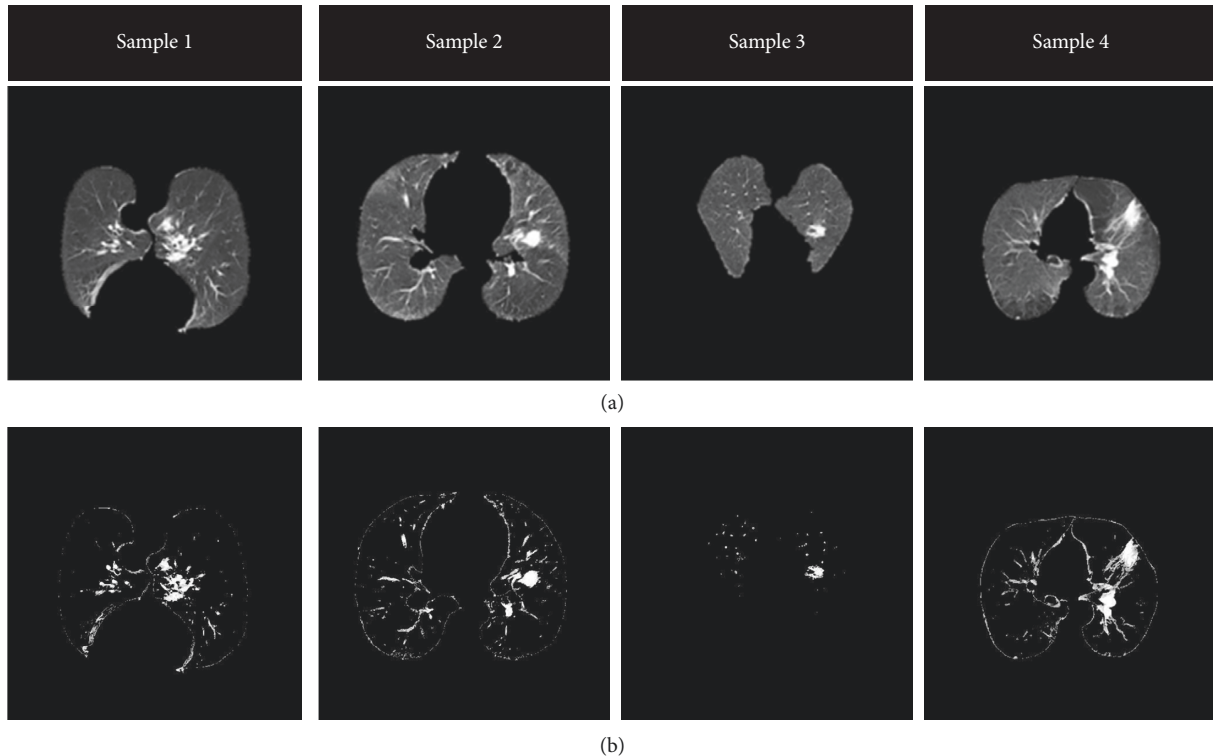
(a)

(b)

FIGURE 4: Examples of segmented lung nodules: (a) images of extracted lung volume and (b) segmentation of lung nodule using SOM.

this study, a total of 16 shape features were extracted to facilitate the determination of nodule type from CT lung images. Statistical features of a gray level image (GTF) of a texture are first derived with the help of GLCM texture features proposed by Haralick [35, 36]. GLCM method shows the relationship between pixels of different gray level and is widely used in applications of medical image processing. In this study, a total of 88 features were extracted with GLCM from 0°, 45°, 90°, and 135° angle directions in $d = 2$ distance. Wavelet decomposition transform can denote distribution of energy features of different regions (TEF). ROI of the CT image is divided into four subbands with 2D wavelet decomposition. Three images are created in low frequencies, and an image is created in high frequencies with wavelet decomposition transform from an image [37]. In this study, 13 energy features of an image are extracted with wavelet decomposition. The number of features extracted by each feature extraction method used in this study is presented in Table 1.

On lung CT images, malign nodules are generally more complex and irregular, while benign nodules are rounder with certain borders. Most of the benign nodules have small variance values. However, malign nodules show relatively higher variance values [3]. Figure 5 shows the examples of benign and malign lung nodules on CT images.

Since 123 features extracted are rather large in size, they may negatively affect accuracy during classification step. Thus, selecting the most appropriate features instead of using all features will be a more efficient method. We used PCA method for dimension reduction of feature vector.

TABLE 1: The number of extracted 2D features from lung CT images.

Feature extraction method	Number of feature	Order
SSF	6	0–6
SBF	16	7–22
GTF	$22 * 4 = 88$	23–110
TEF	13	111–123

PCA is used to reduce dimensionality of large dataset [38, 39]. We can select a number of features only up to one-third of the number of data (patterns) in the smaller of the two classes. Thus, for our work, the smallest class has 104 patterns (benign nodules), and since we split the data to half, one-third of 52 is around 17. Therefore, we selected with PCA the most appropriate 17 features (components) from 123 features. Figure 6 denotes principal component analysis of extracted features with cumulative variance. As can be seen from the chart in Figure 6, it is seen that the variance of the first 20 components is more selective.

3.4. Nodule Classification. In the proposed pipeline, we have used a probabilistic neural network (PNN) model to make automated decision about the nodule types (benign or malignant). PNN is an effective tool for many classification implementations and can easily make classifications [40, 41]. Figure 7 presents the architecture of the PNN designed for this study. Neuron number in the input layer is selected 17 according to the number of inputs.
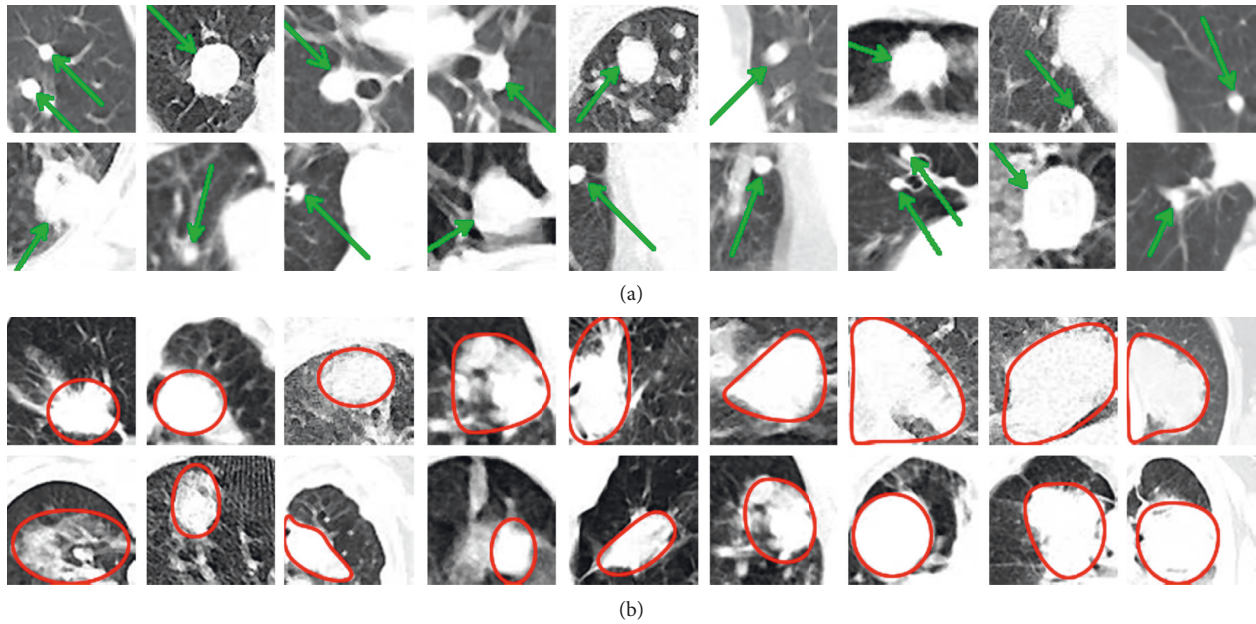
(a)

(b)

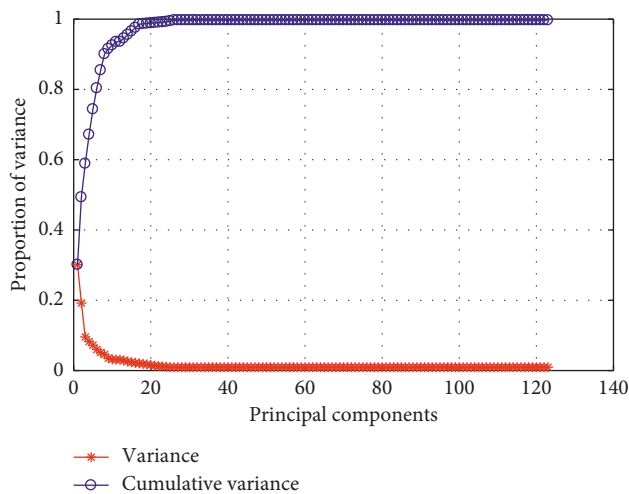Figure 5: Examples of benign (a) and malign (b) lung nodules.

Figure 6: Principal component analysis of extracted features with cumulative variance.

4. Experimental Results

In this work, we realized all experiments using a PC with i7 processor, 16 GB memory, and Windows 10. Moreover, MATLAB software was used for performance evaluation of the proposed pipeline. In all experiments, leave-one-out cross validation technique was run at the level of nodule. So, all of 220 nodules (104 benign and 116 malignant) were used for both trainings and tests. Figure 8 summarizes the processing steps of the proposed pipeline.

4.1. CT Lung Dataset. In this study, an image dataset was prepared for the proposed pipeline. CT examinations were realized by using a helical CT scanner from Sincan Nafiz

Korez Hospital. Its acquisition parameters are slice collimation 1.0 mm and slice width 1 mm. Scans were acquired in 130 kV and 75 mAs. The size of the images was 512×512 pixels. The images were stored as DICOM format files. The database consists of 47 CT scans from 47 different patients. 35 of volunteer patients are male and remaining of them are female. Their ages are between 30 and 79 (mean 58.7 ± 10.5 years). All patients agreed that they have a legal and moral right to accurate and reliable information for the scan. These patients should be given clearly the diagnosis and prognosis with a simple language. There are a total of 9504 CT modality images in the database, and the number of CT slices per scan varies between 116 and 283. After the CT scan, the physician provided the selection of the slice where the nodule is fully visible. 1128 ROI, which includes a total of 220 nodules (104 benign and 116 malignant), were selected from 9504 CT images with the help of a lung physician and three experienced radiologists in the lung parenchyma. This process has been conducted by means of an annotation tool. The nodules were also approved by biopsy. Sizes of nodules change from 3 to 65 mm in diameter. The size distribution of the nodules is shown in Figure 9.

4.2. Validation of LUVEM with Evaluation Metrics. Proposed lung volume extraction method (LUVEM) in this study is compared with the standard manual segmentations using measurement metrics. We evaluate manual segmentations of the expert and automated segmentations of LUVEM using two popular overlap measures. We used a segmentation software tool developed by us for manual segmentation on the dataset. The software tool outlines edges automatically, presenting us to obtain contours of the nodule boundaries. The metrics evaluate the overlapping between the two sets. The first overlap metric, represents the
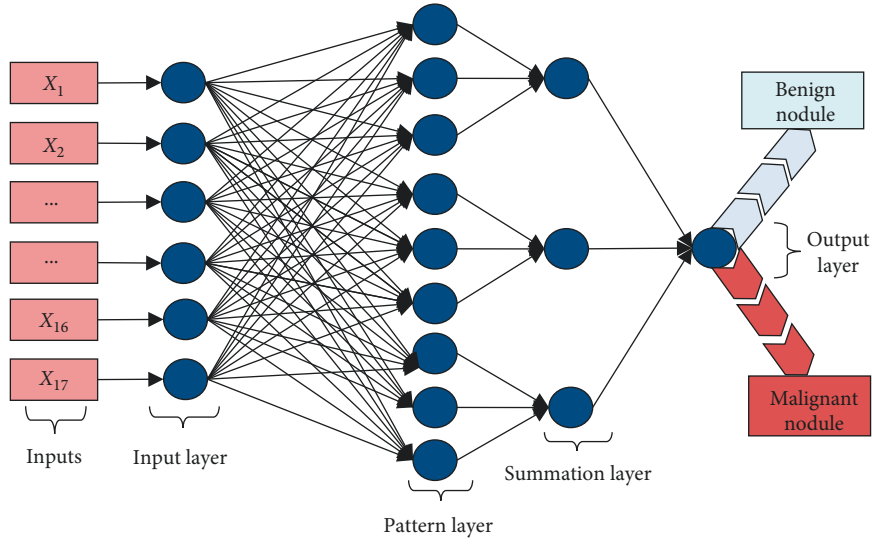
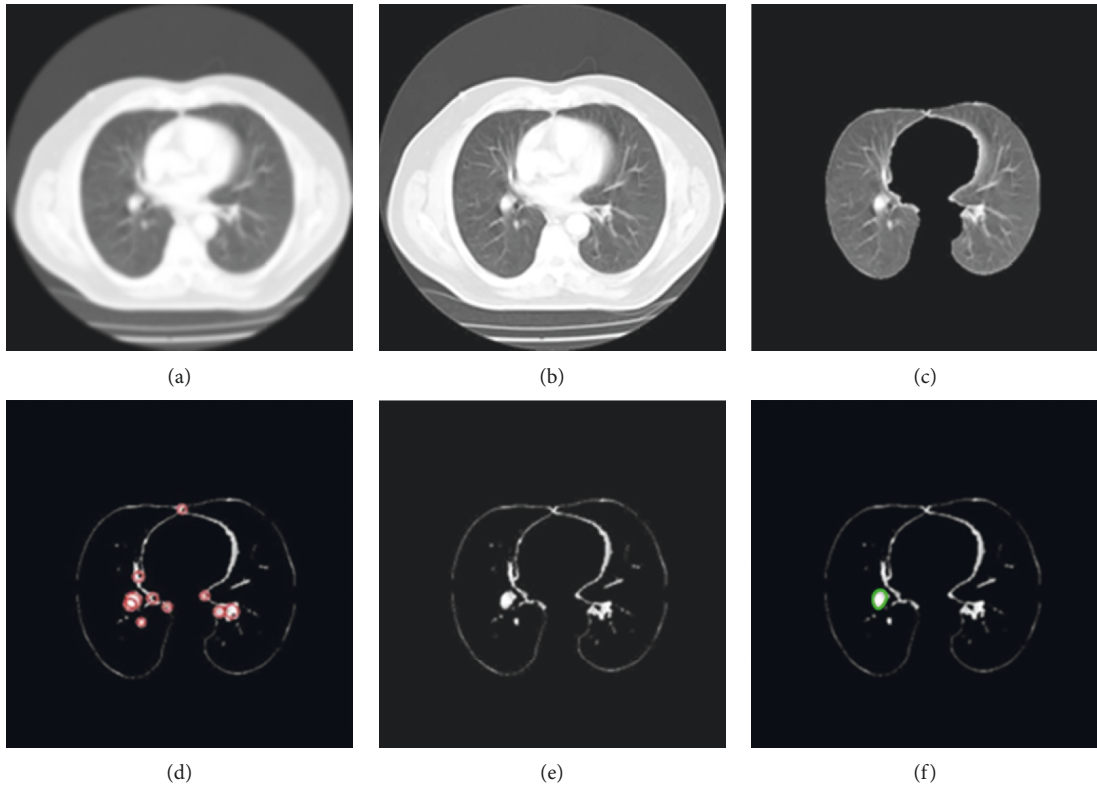FIGURE 7: Probabilistic neural network architecture used in the proposed method for nodule classification.

FIGURE 8: Processing steps of the proposed pipeline: (a) original DICOM image; (b) image preprocessing and enhancement; (c) lung volume extraction from CT scan; (d) detection of candidate nodules; (e) segmentation of nodules; (f) classification of nodules.

Jaccard coefficient (union overlap), defined as intersection over manual and automatic segmentations and measures the similarity of the S1 and S2 sets [42]. Our second overlap metric, the Dice coefficient (mean overlap), gives double the weight to agreements between the two sets [43]. Jaccard and Dice metrics are denoted in the following equations:

$$
\text{Jaccard} = \frac{|S1 \cap S2|}{|S1 \cup S2|},
$$

$$
\text{Dice} = \frac{2|S1 \cap S2|}{|S1| + |S2|}.
$$

(1)
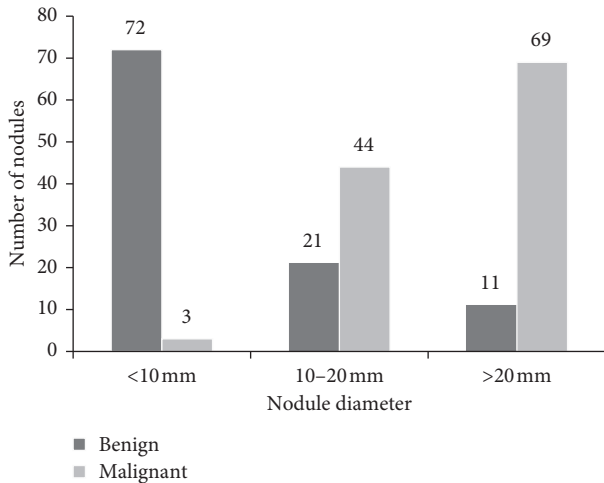
FIGURE 9: Size distribution of benign and malignant nodules in the image dataset.

TABLE 2: The Jaccard and Dice metrics measures for LUVEM.

	Jaccard overlap	Dice overlap
Otsu's method	0.587 ± 0.093	0.786 ± 0.088
LUVEM	**0.867 ± 0.051**	**0.938 ± 0.032**

TABLE 3: Confusion matrixes for feature extraction methods.

FE method	Classification results without PCA				Classification results with PCA			
	TP	FP	FN	TN	TP	FP	FN	TN
SSF	90	22	26	82	90	22	26	82
SBF	105	17	11	87	109	12	7	92
GTF	101	19	15	85	110	11	6	93
TEF	109	15	7	89	111	11	5	93
All FE methods (combined)	111	12	5	92	113	6	3	98

We show the overlap metrics (Jaccard and Dice) that result from both LUVEM and Otsu's method. These results are the comparison of automated segmentations of LUVEM and Otsu methods with manual segmentation on 254 lung CT image in our database. The results in Table 2 showed LUVEM is higher in Jaccard overlap (0.867) and Dice overlap (0.938) than Otsu's method.

4.3. Detection Rates. Confusion matrixes of classification results with PNN according to each feature extraction and PCA method are presented in Table 3. As shown in Table 3, the usage of PCA affects the detection performance of the pipeline positively. Moreover, the usage of combined features extraction methods with PCA gives best success rate.

Table 4 presents the values of performance criteria obtained in the classification results of the proposed pipeline when feature extraction methods were used separately and together. According to the table, performance values are more successful when all feature extraction methods are used together. Accuracy (Acc) was found to be 92.27% when 123 features were used in classification without feature selection through PCA, and this rate was found to be 95.91% with the use of PCA. Similarly, more successful results were obtained in sensitivity (Sen), specificity (Spc), positive decision value (PDV), negative decision value (NDV), and F1 score criteria as presented in Table 4.

Since our CT image database was divided into 3 groups according to the size of nodules such as <10, 10–20, and> 20 mm, we also realized a performance evaluation according to size group of the nodules. These experiments were realized with all together feature extraction method using PCA. Table 5 presents the result of detection performance depending upon nodule size. As shown in Table 5, the proposed pipeline can classify even small nodules with high success rates. Overall detection result of proposed pipeline according to the nodule size is 95.91%.

Receiver operator characteristic (ROC) curve is another popular performance evaluation criteria used in detection

systems. Area under an ROC curve is measured according to sensitivity and specificity values of system. This area shows how the system is successful. Therefore, we also present ROC curve of our proposed detection system. Figure 10 shows ROC curve of the system obtained classification results of each lung nodule group and overall system. As seen in this graphic, area under ROC curve and true positive rate of small size nodules are lower than big size nodules. Here, as can be seen from this figure, if the nodule size is too large and too small, the success rate decreases.

Processing time is another performance criterion that we have used for the evaluation of the proposed pipeline. Longest time is needed for nodule detection step due to the use of SOM method for segmentation. Since SOM is an ANN model, it has a lot of time-consuming mathematical operations. In average, classification of a CT image as benign or malignant takes 2–3 seconds approximately. It can be accepted as a reasonable time period when it is compared with the time it needs for a physician to make decisions.

5. Conclusions and Discussion

In this study, a fully automated pipeline was proposed to classify benign and malign lung nodules on CT images. By means of the designed pipeline, nodule detection as well as benign/malign distinction was performed with high accuracy, sensitivity and specificity rates. Moreover, it was designed a preprocessing method called LUVEM for extracting the lung volume from CT images. SOM method was used to allow successful detection of lung nodules in early stages. According to the detailed experiment performed on large dataset with combined features, the proposed pipeline can differentiate benign/malign nodules with high accuracy rates such as 94.68% (3–10 mm), 96.92% (10–20 mm), and 96.25% (>20 mm) using PNN. The proposed pipeline can be used by the physicians as a supplementary tool for benign and malign nodule classification.

We evaluated the performance of the pipeline on Lung Imaging Database Consortium-Image Database Resource

TABLE 4: Overall performance results of proposed pipeline.

Performance criteria	Classification results without PCA					Classification results with PCA				
	SSF	SBF	GTF	TEF	All	SSF	SBF	GTF	TEF	All
Acc	78.18	87.27	84.55	90.00	92.27	78.18	91.36	92.27	90.00	95.91
Sen	77.57	90.52	87.07	93.97	95.67	77.57	93.97	94.83	95.67	97.42
Spc	78.85	83.65	81.73	85.58	88.46	78.85	88.46	89.42	89.42	94.24
PDV	80.36	86.07	84.17	87.90	90.24	80.36	90.08	90.91	90.98	94.96
NDV	79.93	88.78	85.00	92.71	94.85	79.93	92.93	93.94	94.90	97.03
F1	0.79	0.88	0.85	0.92	0.93	0.79	0.92	0.93	0.94	0.96

TABLE 5: Assessment of performance measurement criteria according to nodule size.

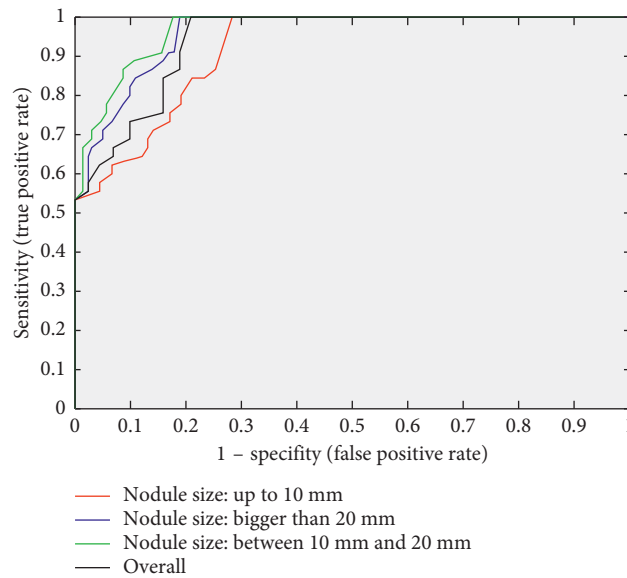| Nodule size (mm)	The number of nodule	Confusion matrix				Performance criteria					
		TP	FP	FN	TN	Acc	Sen	Spc	PDV	NDV	F1
<10	75	3	4	0	68	94.67	100	94.45	42.86	100	0.60
10–20	65	43	1	1	20	96.92	97.73	95.24	97.73	95.24	0.98
>20	80	67	1	2	10	96.25	97.10	90.91	98.53	83.34	0.98
Overall	220	113	6	3	98	95.91	97.42	94.24	94.96	97.03	0.96

FIGURE 10: ROC curve of classification precision in proposed pipeline in different nodule diameter.

TABLE 6: The performance evaluation of proposed pipeline on LIDC-IDRI.

The number of nodules	Confusion matrix				Performance criteria					
	TP	FP	FN	TN	Acc	Sen	Spc	PDV	NDV	F1
38	22	2	4	10	84.21	84.62	83.33	91.67	71.43	0.88

Initiative (LIDC-IDRI) as well [50]. LIDC-IDRI dataset is the largest publicly available reference database for detection of lung nodules. We choose LIDC-IDRI dataset since it contains almost all the related information for lung CT including annotations on nodule sizes, locations, diagnosis results, and other information. We collected a total of 38 lung nodules from the dataset, including 26 malignant and 12 benign nodules. According to the evolution results on the proposed pipeline, accuracy obtained 84.21% using all FE

methods and PCA. In this test, F1 score result was found as 0.88. The obtained performance evaluation values of proposed pipeline on LIDC-IDRI dataset are denoted in Table 6.

There are some advantages of the proposed pipeline compared to the state-of-the-art systems. Firstly, the proposed pipeline has two diagnosis possibilities. It can perform nodule detection together with nodule classification. Second advantage of the proposed pipeline is to provide the detection of small nodules in the lung with the use of SOM

TABLE 7: The comparison of our pipeline with previously published CADs.

CAD system	CT image database	Number of cases	Nodule size(mm)	Sensitivity (%)	Average FPR
Dehmenski et al. [9]	Their own database	70	3–20	90.0	14.6
Suarez-Cuenca et al. [10]	Their own database	22	4–27	80.0	7.7
Opfer and Wiemeker [46]	LIDC database [47, 48, 50]	93	≥4	74.0	4
Rubin et al. [51]	Their own database	20	≥3	76	3
Sahiner et al. [49]	LIDC database [47, 48, 50]	48	3–36.4	79	4.9
Messay et al. [24]	LIDC database [47, 48, 50]	84	3–30	82.66	3
Suzuki et al. [52]	Their own database	101	8–20	80.3	16.1
Park et al. [53]	Their own database	38	Indefinite	80	–
Choi and Choi [23]	LIDC database [47, 48, 50]	32	3–30	94.1	5.45
Choi and Choi [44]	LIDC database [47, 48, 50]	58	3–30	95.28	2.27
Proposed method	Our database	47	3–35	97.42	4.54

method during segmentation step. This is remarkable in terms of early detection of lung cancer. Third advantage of the proposed pipeline is to have relatively high detection performance. Accuracy, sensitivity, and specificity of the system were calculated as 95.91%, 97.42%, and 94.24%, respectively. It is fairly difficult to compare formerly reported CAD systems due to different datasets, nodule types, sizes, and validation methods. We picked out some CAD systems to compare their performances. Some of them [23, 24, 44–46] used the LIDC database [47–49], and the other used their own databases. Table 7 denotes the comparison of the proposed pipeline with some CAD systems. When the results are analyzed, our pipeline has high sensitivity on our CT image dataset.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.
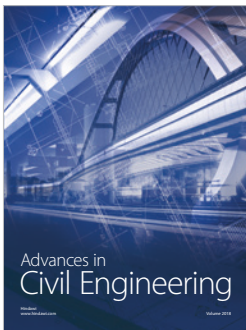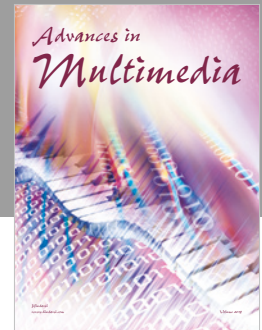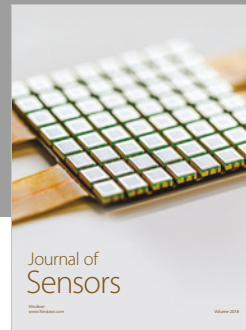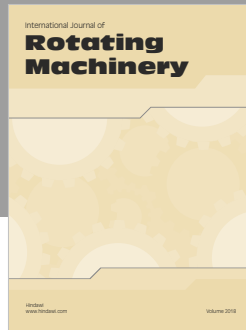
Acknowledgments

References

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 62, no. 1, pp. 10–29, 2012.

[2] C. Jacobs, E. M. Van Rikxoort, T. Twellmann et al., "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis*, vol. 18, no. 2, pp. 374–84, 2014.

[3] E. Dandil, M. Cakiroglu, Z. Eksi, M. Ozkan, O. K. Kurt, and A. Canan, "Artificial neural network-based classification system for lung nodules on computed tomography scans," in *Proceedings of 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 382–86, Tunis, Tunisia, August 2014.

[4] N. Howlader, A. Noone, M. Krapcho et al., *SEER Cancer Statistics Review, 1975-2011*, National Cancer Institute, Bethesda, MD, USA, 2014.

[5] 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 Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 648–659, 2015.

[6] Y. J. Jeong, C. A. Yi, and K. S. Lee, "Solitary pulmonary nodules: detection, characterization, and guidance for further diagnostic workup and treatment," *American Journal of Roentgenology*, vol. 188, no. 1, pp. 57–68, 2007.

[7] S. Ozekes and A. Y. Camurcu, *Automatic Lung Nodule Detection Using Template Matching*, Springer, Berlin, Germany, 2006.

[8] A. M. Schilham, B. Van Ginneken, and M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database," *Medical Image Analysis*, vol. 10, no. 2, pp. 247–58, 2006.

[9] J. Dehmeshki, X. Ye, X. Lin, M. Valdivieso, and H. Amin, "Automated detection of lung nodules in CT images using shape-based genetic algorithm," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 408–17, 2007.

[10] J. J. Suarez-Cuenca, P. G. Tahoces, M. Souto et al., "Application of the iris filter for automatic detection of pulmonary nodules on computed tomography images," *Computers in Biology and Medicin*, vol. 39, no. 10, pp. 921–33, 2009.

[11] K. Murphy, B. Van Ginneken, A. M. Schilham, B. J. De Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical Image Analysis*, vol. 13, no. 5, pp. 757–70, 2009.

[12] M. L. Giger, K. T. Bae, and H. Macmahon, "Computerized detection of pulmonary nodules in computed tomography images," *Investigative Radiology*, vol. 29, no. 4, pp. 459–65, 1994.

[13] J. Hasegawa, K. Mori, J. Toriwaki, H. Anno, and K. Katada, "Automated extraction of lung cancer lesions from multi-slice chest CT images by using three-dimensional image processing," *Systems and Computers in Japan*, vol. 25, no. 11, pp. 68–77, 1994.

[14] K. Kanazawa, Y. Kawata, N. Niki et al., "Computer-aided diagnosis for pulmonary nodules based on helical CT images," *Computerized Medical Imaging and Graphics*, vol. 22, no. 2, pp. 157–67, 1998.

[15] K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and

malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1138–50, 2005.

[16] T. Sun, J. Wang, X. Li et al., "Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 2, pp. 519–24, 2013.

[17] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202–9, 2014.

[18] M. Javaid, M. Javid, M. Z. U. Rehman, and S. I. A. Shah, "A novel approach to CAD system for the detection of lung nodules in CT images," *Computer Methods and Programs in Biomedicine*, vol. 135, pp. 125–139, 2016.

[19] M. Z. ur Rehman, M. Javaid, S. I. A. Shah, S. O. Gilani, M. Jamil, and S. I. Butt, "An appraisal of nodules detection techniques for lung cancer in CT images," *Biomedical Signal Processing and Control*, vol. 41, pp. 140–151, 2018.

[20] Z. Wang, J. Xin, P. Sun, Z. Lin, Y. Yao, and X. Gao, "Improved lung nodule diagnosis accuracy using lung CT images with uncertain class," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 197–209, 2018.

[21] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109–119, 2019.

[22] DICOM (Digital Imaging and Communications in Medicine), http://medical.nema.org.

[23] W. J. Choi and T. S. Choi, "Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images," *Information Sciences*, vol. 212, pp. 57–78, 2012.

[24] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Medical Image Analysis*, vol. 14, no. 3, pp. 390–406, 2010.

[25] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 7, pp. 1810–1820, 2009.

[26] S. Akram, M. Y. Javed, A. Hussain, F. Riaz, and M. Usman Akram, "Intensity-based statistical features for classification of lungs CT scan nodules using artificial intelligence techniques," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 27, no. 6, pp. 1–15, 2015.

[27] R. O. Duda and P. E. Hart, "Use of Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[28] H. Yazid, H. Yazid, M. Harun et al., "Circular discontinuities detection in welded joints using Circular Hough Transform," *NDT and E International*, vol. 40, no. 8, pp. 594–601, 2007.

[29] T. Kohonen, "Self-organizing maps of massive databases," *Engineering Intelligent Systems for Electrical Engineering and Communications*, vol. 9, no. 4, pp. 179–185, 2001.

[30] S. Haykin, *Neural Networks*, A Comprehensive Foundation, Hamilton, ON, Canada, 1999.

[31] D. Chi, "Self-organizing map-based color image segmentation with k-means clustering and saliency map," *ISRN Signal Processing*, vol. 2011, Article ID 393891, 18 pages, 2011.

[32] C. Hollitt, "A convolution approach to the circle Hough transform for arbitrary radius," *Machine Vision and Applications*, vol. 24, no. 4, pp. 683–694, 2013.

[33] U. Akilandeswari, R. Nithya, and B. Santhi, "Review on feature extraction methods in pattern classification," *European Journal of Scientific Research*, vol. 71, no. 2, pp. 265–272, 2012.

[34] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," *Pattern Recognition*, pp. 43–90, 2008.

[35] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[36] R. M. Haralick, K. Shanmuga, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[37] G. Van, P. Wouver, D. Scheunders, and D. Van, "Statistical texture characterization from discrete wavelet representation," *IEEE Transactions on Image Processing*, vol. 8, no. 4, pp. 592–598, 1998.

[38] H. Camdevyren, N. Demyr, A. Kanik, and S. Keskyn, "Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs," *Ecological Modelling*, vol. 181, no. 4, pp. 581–589, 2005.

[39] L. H. Chen and S. Y. Chang, "An adaptive learning algorithm for principal component analysis," *IEEE Transactions on Neural Networks*, vol. 6, no. 5, pp. 1255–1263, 1995.

[40] K. Z. Mao, K. C. Tan, and W. Ser, "Probabilistic neural-network structure determination for pattern classification," *IEEE Transactions on Neural Networks*, vol. 11, no. 4, pp. 1009–1016, 2000.

[41] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[42] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[43] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[44] W.-J. Choi and T.-S. Choi, "Automated pulmonary nodule detection system in computed tomography images: a hierarchical block classification approach," *Entropy*, vol. 15, no. 2, pp. 507–523, 2013.

[45] J. J. Erasmus, J. E. Connolly, H. P. Mcadams, and V. L. Roggli, "Solitary pulmonary nodules: Part I. Morphologic evaluation for differentiation of benign and malignant lesions 1," *Radiographics*, vol. 20, no. 1, pp. 43–58, 2000.

[46] R. Opfer and R. Wiemker, "Performance analysis for computer aided lung nodule detection on LIDC data,"art. no. 65151C, *Medical Imaging 2007: Image Perception, Observer Performance, and Technology* Assessment, vol. 6515, pp. C5151–C5151, 2007.

[47] M. F. Mcnitt-Gray, S. G. Armato, C. R. Meyer et al., "The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation," *Academic Radiology*, vol. 14, no. 12, pp. 1464–1474, 2007.

[48] A. P. Reeves, A. M. Biancardi, T. V. Apanasovich et al., "The Lung Image Database Consortium (LIDC): a comparison of different size metrics for pulmonary nodule measurements," *Academic Radiology*, vol. 14, no. 12, pp. 1475–1485, 2007.

[49] B. Sahiner, L. M. Hadjiiski, H. P. Chan et al., "Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: observer performance study,"art. no. 65151D, *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*, vol. 6515, pp. D5151–D5151, 2007.

[50] S. G. Armato III, G. Mclennan, M. F. Mcnitt-Gray et al., "Lung image database consortium: developing a resource for the medical imaging research community 1," *Radiology*, vol. 232, no. 3, pp. 739–748, 2004.

[51] G. D. Rubin, J. K. Lyo, D. S. Paik et al., "Pulmonary nodules on multi–detector row CT scans: performance comparison of radiologists and computer-aided detection 1," *Radiology*, vol. 234, no. 1, pp. 274–283, 2005.

[52] K. Suzuki, S. G. Armato III, F. Li, S. Sone, and K. Doi, "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography," *Medical Physics*, vol. 30, no. 7, pp. 1602–1617, 2003.

[53] S. C. Park, J. Tan, X. Wang et al., "Computer-aided detection of early interstitial lung diseases using low-dose CT images," *Physics in Medicine and Biology*, vol. 56, no. 4, pp. 1139–53, 2011.