ACCSCIENCE PUBLISHING

Tumor Discovery

ORIGINAL RESEARCH ARTICLE

An approach for classification of lung nodules

Naveen HM[1]*, Naveena C[1], and Manjunath Aradhya VN[2]

[1]Department of Computer Science Engineering, SJB Institute of Technology, Bangalore, Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

[2]Department of Master Computer Application, JSSTU, Mysuru, Affiliated to JSS Science and Technology University, Mysuru, Karnataka, India

Abstract

The main objective of the proposed work is to develop an automated computer-aided detection (CAD) system to classify lung nodules using various classifiers from computed tomography (CT) images. One of the most important steps in lung nodule detection is the classification of nodule and non-nodule patterns in CT. The early detection of the condition helps lower the mortality rate. The developed CAD systems consist of segmentation, feature extraction, and classification. In this work, a filter method is used to segment the infected region. Later, we extracted features through and fed into classifiers such as Decision Stump (DS), Random Forest (RF), and Back Propagation Neural Network (BPNN). The experimentation was conducted on LIDC-IDRI dataset, and the results with BPNN outperformed those with DS and RF classifiers.

Keywords: Decision stump; Random forest; AdaBoost-Decision stump; AdaBoost-Random forest; Back propagation neural network

***Corresponding author:**
Naveen HM
(naveenhm056@gmail.com)

Citation: Naveen HM, Naveena C, and Aradhya VNM, 2023, An approach for classification of lung nodules. *Tumor Discov,* 2(1): 317. https://doi.org/10.36922/td.317

Received: December 28, 2022
Accepted: February 17, 2023
Published Online: March 8, 2023

1. Introduction

The second most frequent cancer in both men and women is thought to be lung cancer. It is the main factor in cancer-related fatalities. According to the most recent estimates, there are around 7.6 million cancer-related deaths globally each year, according to the most recent numbers supplied by the World Health Organization[1]. Furthermore, it is anticipated that the number of deaths from lung cancer would keep increasing, reaching almost 17 million in 2030. Successful treatment of lung cancer depends greatly on early detection. Significant data suggest that early identification of lung cancer will reduce mortality rates[2]. Lung cancer in an early stage manifests itself as a pulmonary nodule, which grows rapidly and later becomes a tumor. The characteristics of pulmonary nodules are based on calcification, internal structure, sphericity, speculation, subtlety, and texture. Nodules usually appear smaller in medical images. Hence, detection of pulmonary nodule is one of the most challenging tasks[3].

Various imaging techniques, including radiography, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography-CT (PET-CT), among others, can be used to detect pulmonary nodules. Radiologists face a challenging problem when trying to find lung nodules on radiographs, because nodules present behind the rib cages are hidden and the miss rate could increase up to 30%[4,5]. MRI and PET-CT techniques are more expensive and time-consuming. The

CT imaging is less expensive and produces a variety of cross-sectional images of complete chest within a single breath hold. For the analysis and early detection of lung nodules, it is now regarded as the best imaging technique. Majority of the pulmonary nodules are benign; however, a small populace of them grow to be malignant. The radiologists examine the CT scan to conclude whether a nodule presents a chance for malignancy. The radiologist finds it challenging and time-consuming to detect some nodules in the CT because of non-pathological features[6]. To get around this, radiologists opt for a computer-assisted approach as a backup method to validate their interpretation. Computer-based processes like computer-aided detection (CAD) act as the radiologist's "second pair of eyes" to analyze medical pictures for any problematic regions. CAD is a topical technique designed to improve the radiologists' ability to find even the smallest lung nodules at their earliest stages.

The main objective of CAD is to improve disease identification by lowering the false negative (FN) rate brought on by observational omission. CAD was created with the purpose of improving systematic clinical decision-making and the performance of detection in medical imaging modalities. The aim of this paper is to focus on the architecture of different stages of CAD design with fruitful results to assist the radiologists in detecting lung nodule at early stage.

2. Literature survey

In the field of lung cancer and in relation to the work of this study, there are a number of existing models and algorithms. The stages of CAD systems can be used to categorize the existing work.

Various techniques, such as region growing[7,8], watershed segmentation[9], fuzzy logic, active contours, intensity-based thresholding, graph search algorithm[10], etc., were used for region of interest (ROI) image segmentation Segmentation is an essential step in nodule detection. The survey in various existing segmentation techniques for lung nodule is discussed in detail. Dai *et al.*[11] developed a segmentation algorithm known as segmentation by registration. The authors compared their work with algorithms such as automatic region growing, interactive region growing and voxel classification. Daneshmand *et al.*[12] demonstrated a precise technique designed for toning lung nodules in chest CT scans. The region growing, optimal thresholding and optimal cube registration were used in this system Dawoud *et al.*[13] proposed an adaptive border marching which is a geometric-based algorithm used for segmenting the lung. De Nunzio *et al.*[14] performed phantom experiment for CT image. The density of the nodule and its size was measured using the point spread function. Deep *et al.*[15] presented an algorithm for segmenting different nodule types such as juxtravascular, pleura tail, juxtapleural, solid, and non-solid nodules.

Segmentation technique such as region growing and some hybrid fuzzy connectivity models was implemented. Another segmentation scheme proposed by Dehmeshki *et al.*[16] used two different datasets taken from the LIDC database. Two different techniques such as dynamic programming model and multidirection fusion techniques are used to know the information, relationship between adjacent slices and to reduce segmentation error. Dehmeshki *et al.*[17] utilized a three-step procedure to find lung nodules. Initially, lung regions are segmented using an adaptive threshold algorithm. Second, lung vessel was removed using active contour model (ACM) and finally, the suspicious nodules were located using a Hessian matrix (selected shape filter). Delogu *et al.*[18] developed a fully automated segmentation method for detecting the pulmonary nodule. The authors used region growing approach for a set of 130 CT images. Only 84 images produced satisfactory result. Dheepak *et al.*[19] proposed a methodology for segmentation of juxtapleural lung nodules. The authors used two techniques for detecting the nodules from the lung. They are region growing and shape curvature-based techniques. Doi *et al.*[20] developed a method for segmenting the lung with the help of CT data. The method was fully automatic and was composed of two steps. They are robust active shape model (RASM) and optimal surface finding method. Dolejsi *et al.*[21] proposed a CAD to reduce the lung volume and juxtapleural nodule from thoracic CT images. For segmenting the lung volume and nodule in juxtapleural, region growing and a 3D-mass-spring mdel (MSM) was used. Elizabeth *et al.*[22] proposed a region-based ACM based on local divergence energies. This model was designed for blurred boundary and noisy images. The author used regularization function to smoothen the boundary from different noise level. The system performance was evaluated with Chan-Vese's (CV) model, region scalable fitting and local Gaussian fitting. Enquobahrie *et al.*[23] developed an edge detection model, which precisely diffuses the edge space. Farag *et al.*[24] proposed a newly developed nodule segmentation algorithm which was stable, accurate, and automated. Farag *et al.*[25] developed a segmentation model to segment the juxtavascular and ground glass (GG) nodules. The authors proposed parametric mixture model for juxtavascular nodules and ACM for detecting leakage boundary.

A nonlinear level set method proposed by Farag *et al.*[26] used adaptive velocity function and edge stopping function to employ a noise-free segmentation model. Gambhir *et al.*[27] focused on segmenting the lung nodule

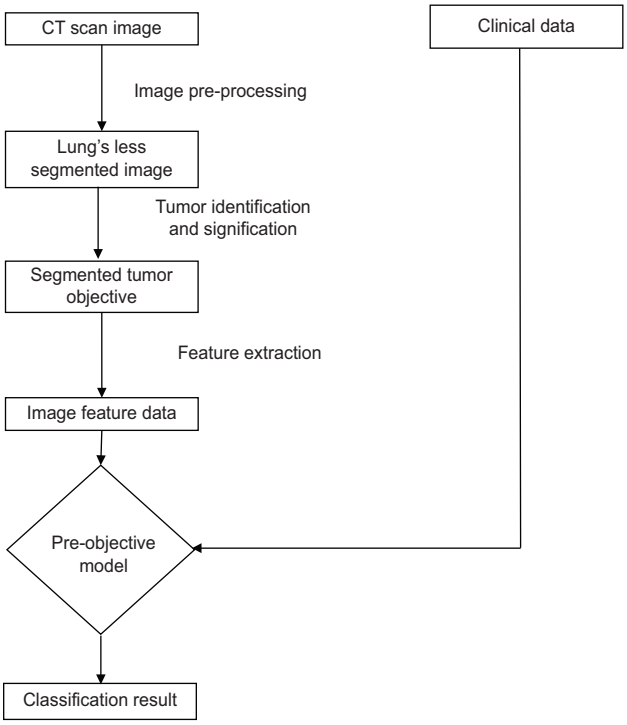

Figure 1. General workflow of the process of developing and using predictive models.

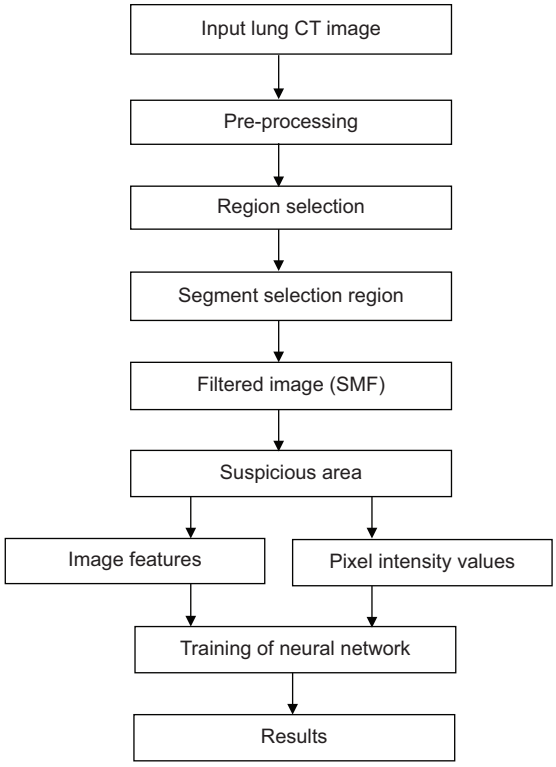

Figure 2. Pre-processing selective median filter.

with lesser number of false positive (FP) findings. Garro *et al.*[28] proposed a method to segment the juxtapleural nodule and lung vessels from the CT image. Golosio *et al.*[29] segmented the pleural and vessels in lung CT. Gomathi *et al.*[30] developed a segmentation method to improve nodule detection accuracy. The authors mainly focused on juxtapleural nodule for image segmentation. A parameter-free algorithm such as bidirectional chain coding method was used to smoothen the lung border. Gomathi *et al.*[31] presented a segmentation algorithm to produce efficient and accurate result. An improved graph that cuts algorithm along with Gaussian mixture models (GMMs) was proposed to segment the lung nodule. Gon alves *et al.*[32] developed a hybrid segmentation technique which combined the fully automatic and semi-automatic global segmentation technique. Gould *et al.*[33] formulated central medialness adaptive principle, a Hessian-based strategy, to segment the lung nodule in CT images. Multi-resolution contour let transform Grigorescu *et al.*[34] can also be used to extract the features. These features are used for further processing in the classification, which is the final stage of the CAD system. Gu *et al.*[35] proposed a technique to detect the nodule using template-based model. The minimum and maximum Hounsfield density (HU) was obtained from the intensity of nodule data. Shape-based or shape-texture-based methods resulted in an overall detection process with the lowest accuracy. The existing segmentation techniques produced low accuracy, high error rate, reduced similarity coefficient, long computation time, etc. Medical image segmentation is difficult due to complexity and diversity of anatomical structures on one hand and particular properties such as noise and low contrast (non-solid nodules), on the other hand.

Gudise *et al.*[36] comparative study is made on the computational requirements of the PSO and BP as training algorithms for neural networks. Hua *et al.*[37] presents an automatic algorithm for pathological lung CT image segmentation that uses a graph search driven by a cost function combining the intensity, gradient, boundary smoothness, and the rib information. Jacobs *et al.*[38], a CAD system that combines the output of two prototype CAD systems aimed at detection of ground glass nodules and solid nodules, respectively, could lead to efficient detection of the entire spectrum of lung nodules in chest CT scans.

Shen *et al.*[39] proposes a parameter-free lung segmentation algorithm with the aim of improving lung nodule detection accuracy, focusing on juxtapleural nodules. A bidirectional chain coding method combined with a support vector machine (SVM) classifier is used to selectively smooth the lung border while minimizing the over-segmentation of adjacent regions. Shen *et al.*[40] proposed a robust segmentation technique based on an extension to the traditional fuzzy c-means (FCM) clustering algorithm.

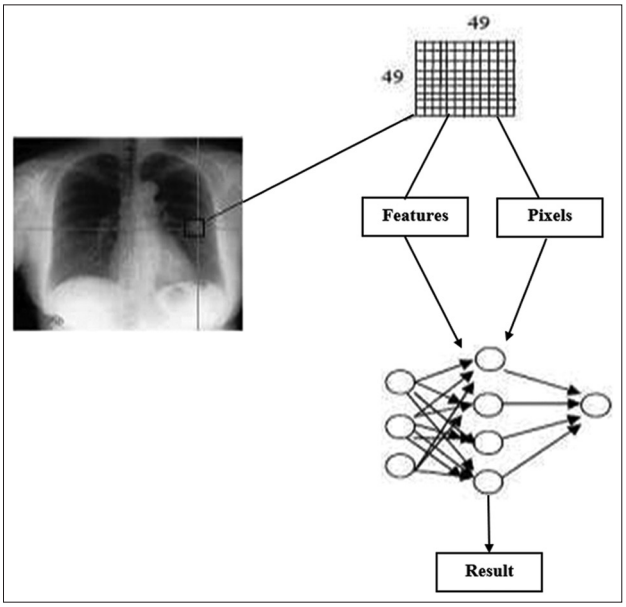

Figure 3. Layer-based region segmentation.

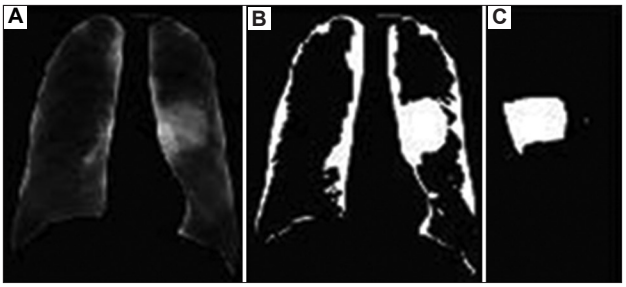

Figure 4. (A) Image after growing, (B) thresholding, and (C) segmented nodule.

Shi *et al.*[41] presents the Optimized Kalman Particle Swarm (OKPS) filter. This filter results from two years of research and improves the Swarm Particle Filter (SPF). Shih-Chung *et al.*[42] presented to predict long term survival versus short term survival. Forty adenocarcinoma diagnostic lung computed tomography (CT) scans from Moffitt Cancer Center were analyzed for survival prediction. A decision tree classifier was able to predict the survival group with an accuracy of 77.5%.

Yuan[43] proposed model can handle blurry boundaries and noise problems. In addition, the regularity of the level set function is intrinsically preserved by the level set regularization term to ensure accurate computation. Zhou S *et al.*[44] proposed a fast and fully automatic scheme based on iterative weighted averaging and adaptive curvature threshold is proposed in this study to facilitate accurate lung segmentation for inclusion of juxtapleural nodules and pulmonary vessels and ensure the smoothness of the lung boundary.

In this work, we proposed efficient method for segmentation and classification of lung nodules.

3. Geometric features

Geometric features are considered the first set of features. The vital structural information of tumor can be easily analyzed with the 2D and 3D geometric features. The evaluation of the geometric features is very useful in quantifying and analyzing the biomedical images like CT scans (2D, 3D)[45].

Image object is formed by the numerous pixels and is rescaled using unit information. The one unit is the area of the single pixel, which denotes that the number of pixels forming the image is the area of the image. If we have the unit information of the image data provided, then the area of the whole image is equal to the product of the area covered by one pixel and the number of pixels unit in the image object. In this chapter, a fully automatic method is described by the authors to detect the cancer in the lungs. This method comprises three sequential steps. The first step is to implement the gray level thresholding method to separate lung region from the image. The second step is the detection of the anterior and posterior junctions to separate the left lung and right lung region. The final step is the smoothening on the boundary of lung along the mediastinum. There are some differences between our proposed and the previous works.

The authors demonstrated that an automated texture mapping methods. The proposed work is experimental in nature: we propose an efficient technique to discover the gray scale qualities of an HRCT dataset with the co-training paradigm. We utilize an effective technique to enhance classifiers that are prepared with not very many posterior and anterior intersection lines between the marked illustrations utilizing a huge pool of concealed right and left lungs.

Finally, to get more cases, there are two or more disjoint functions called views. Processing time and stable results even leaving the lung. It has also been shown that the structures named by experts are smooth with the lung and can be connected stepwise within the frame with irregular boundaries along the mediastinal pathway (Kawane *et al.*[46]). The outcomes are likewise analyzed against "density mask," as of now a standard approach utilized for emphysema recognition in medicinal picture analysis and other automated procedures utilized for arrangement of emphysema in the literature. The new framework can group diffuse districts of emphysema beginning from a bullous setting.

The classifiers worked at various iterations additionally seem to demonstrate an intriguing relationship with

various levels of emphysema, which merits more investigation[47]. The authors suggested using form features to distinguish between obstructive lung infections, and the results showed increased classification sensitivity when compared to features based solely on texture. Be that as it may, their proposed framework is reliant on the region size, for example, 16 × 16, 32 × 32, and 64 × 64 pixels. Gathering region pictures from CT picture is not a simple errand particularly with settled size of area. To expand the proficiency while safeguarding the high sensitivity, another component is wanted[48]. In this paper, a novel feature known as continuous local histogram (CLH) is presented. CLH coordinates three fundamental sorts of features, which are brightness, texture feature, and shape feature, to build the separation.

4. Proposed method

The proposed method is the new dynamic multi-level CAD framework to automatically identify the defects in the lung CT image. This work also employed the enhanced selective median filter (SMF) to increase the quality of the image with clear view and noise reduction. A new neural network multi-level classifier segmentation method was used on the quality-improved CT image to eliminate the suspicious region. The last process classification was done on the CT scan images using the new neural network-based multi-level classifier with the extracted textural features as shown in Figure 1.

The intention of the extraction is to discover the tumor area on the lung area. In some tumor area with tissues and after the binarization, the tumor area would not be included in the lung area. The image morphology can be used to reduce image quality. This paper demonstrated an innovative multi-level boundary repair procedure to enhance the lung CT images using SMF for enhancing the image quality. The phases of the process are as follows: utilizing the less and more circumstances of erosion and expansion operation with respect to the real images. We can just utilize the less circumstances of erosion and operations to achieve the less time of testing.

5. Implementation steps

The initial process will always be the image processing routine of separation of lung from the other internal organs on the chest CT images and finding the area suspected for the presence of nodule. During first stage, the method extracts square areas of 32 × 32 with the suspicious part at the center. Since it is a multi-level region-based and pixel-based technique the inputs to the system are in the square. The pixel's intensity values falling within the suspicious region that are separated and stored in a database will be used in training the system at the next stages. The training of a new neural network multi-level segmentation process is the second stage. It depends on the two types of input, namely, dynamic feature-based inputs and pixel-based inputs. In dynamic feature-based inputs, the first and second order dynamic features are taken into account. In pixel-based inputs, the pixel's intensity values are used in detecting suspicious region. A SMF is used in pre-processing to enhance the quality of the image by enhancing the poor contrast due to noise and effect due to poor lightning conditions while capturing the image and glare. The generation of low-frequency image is done by placing the median pixel value in every pixel value location. The median value of the pixel is calculated on the square area of 8 × 8 pixels centered at the pixel location. The methods used for enhancing the contrast of the images are sharpening and histogram equalization as shown in Figure 2.

6. Lung region segmentation

Due to the active shape models' availability in the database, lung masks were constructed using them. When segmenting the lung region in CT scans, the user can locate the scope by choosing the questionable locations.

For the purpose of selecting the suspicious zone, a 49 × 49 square mask was created using nodules with a resolution of 96 pixels per inch and a diameter of 13 mm. The image database contains nodules with sizes ranging from 8.9 mm to 29.1 mm, with an average of 17.4 mm. The lung nodule is considered to be between the size of 5mm and 20 mm, and it is detected at the initial stage. The input patterns for the classification stage come from the feature vector from the extraction procedure and the selection stage. The dataset on lung CT images were utilized in training the classifiers and for performance evaluation as shown in Figure 3. Basically, region growing is the conceptually better and the simplest approach for image processing. The segmented region is formed by combining the pixel units of same intensity values in this algorithm. A couple of quantized pixels of the same amplitude will be paired together to form a group called atomic region in the initial phase of the process. The process of combining the weak and combining boundaries between the regions is done in the second evaluation as shown in Figure 4.

Table 1. Geometrical features

Features	Value
Area	2815
Perimeter	226.85
Diameter	59.686
Irregularity index	0.69

6.1. Lung nodule segmentation

In region growing, the procedure called labeling is done to put the negative number representing the label of the region to which the pixel has a place. The labeling process keeps tracks of the record of pixels which are yet to be replaced by the labeling. With this reference list, the operation of insertion and removal will be carried out. Removal is done by eliminating the pixels from the front list, and insertion is by inserting the pixels at the end of the list.

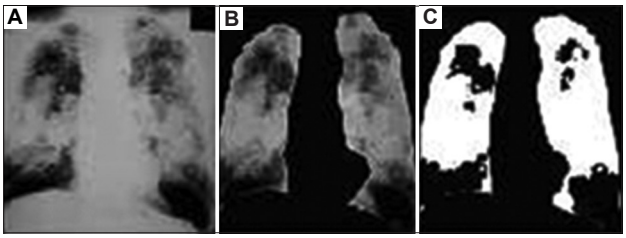

Figure 5. (A) Original image, (B) separated lung fields, and (C) separated cancerous portion

Table 2. Calculated SMF of the filtered images for all types of filters

Filter type	Output SMF
Average filter	7.3133
Weighted average filter	4.2478
Gaussian filter	6.4443
Selective median filter	8.5816
Wavelet filter	8.8732
Wiener filter	13.3969

SMF: Selective median filter

7. Feature extraction

The diameter, area, irregularity index, and perimeters are some of the geometrical features that will be evaluated from the separated lung nodules. The number of pixels in the picture array with the value 1 will be used to determine the size of the segmented tumor image. The area is estimated using the established technique using bit quads, or 2-by-2pixel patterns. The quantity of boundary pixels provides an estimate of the tumor image's perimeter. In terms of its morphology, the tumor has a circular form. The circulatory index will be calculated to identify the irregularity in the circular shape using the expression $I = 4\pi A/p2$, where A is the area of the tumor and P is the perimeter of the tumor in pixels. Table 2 shows the geometrical features for Figure 5. The important features used in the lung cancer classification are texture or contrast features.

The 1st order statistic and 2nd order statistic are the categories under which the contrast features are classified. The shorter processing time and lower minimum cost are the advantages in feature extraction using wavelet due to the solid depiction of the wavelet transform.

8. Classification of lung nodule

Classification is the process of determining whether the nodules belong to a specified class or not. In supervised classification, samples and anticipated classes are known before a classifier is trained on a set of data. It is evident that supervised classification-based methods consistently outperform other types of traditional classification methods[43]. Classification of nodule in the lung is accomplished using

Table 3. Performance metrics of solid feature for various inertia weights

Classifi ers	Accuracy	Sensiti vity	Specificit y	PPV	NPV	F-measu re	G-mean	MC
DS	88	92	84	85.18	91.30	87.81	87.9	0.76
RF	88	92	84	85.18	91.30	87.81	87.9	0.76
Ada-DS	86	92	80	82.14	90.9	85.58	85.79	0.72
Ada-RF	90	100	80	83.33	100	88.89	89.44	0.81
BPNN	98	96	100	100	96.15	97.95	97.97	0.96

PPV: Positive predictive value, NPV: Negative predictive value, DS: Decision Stump, RF: Random Forest, BPNN: Back Propagation Neural Network

Table 4. Performance metrics of part-solid feature for various inertia weights

Classifie rs	Accura cy	Sensitiv ity	Specifi city	PPV	NPV	F-measur e	G-mean	C
DS	77.6	60.19	95.14	92.53	70.5	73.54	76.17	0.59
RF	86.2	85.4	86.4	86.27	85.57	85.49	85.49	0.72
Ada-DS	87.9	83.5	92.2	91.48	84.82	86.81	87.74	0.61
Ada-RF	86.4	85.4	87.4	91.48	84.82	87.89	86.39	0.76
BPNN	93.68	91.26	96.1	95.91	96.11	93.43	93.46	0.87

PPV: Positive predictive value, NPV: Negative predictive value, DS: Decision Stumpe, RF: Random Forest, BPNN: Back Propagation Neural Network

Table 5. Performance metrics of non-solid feature for various inertia weights

Classifier s	Accura cy	Sensitiv ity	Specific ity	PPV	NPV	F-measur e	G-mean	MC
DS	77.8	100	55.6	69.23	100	71.46	74.56	0.59
RF	91.7	83.3	100	100	85.71	90.88	91.26	0.83
Ada-DS	94.4	88.9	100	100	90	94.12	94.28	0.88
Ada-RF	94.4	100	88.9	88.88	100	94.12	94.28	0.88
BPNN	97.2	100	88.9	94.73	100	97.11	97.15	0.94

PPV: Positive predictive value, NPV: Negative predictive value, DS: Decision Stump, RF: Random Forest, BPNN: Back Propagation Neural Network

different classifiers in this work. Classifiers such as Decision Stump (DS), Random Forest (RF), AdaBoost-Decision Stump, AdaBoost-Random Forest, and Back Propagation Neural Network (BPNN) are used. Classifiers are trained to distinguish the true nodule from false nodule.

8.1. Confusion matrix

A confusion matrix is a chart that is used to describe the classifier's performance on a set of predicted condition for which the actual conditions are known. If the radiologist identifies a patient as disease present and the proposed CAD indicates the presence of disease, the detection test result would yield true positive (TP); if the radiologist identifies a patient as nodule absent and the proposed CAD indicates the presence of nodule, the detection test result would yield FP; if the radiologist identifies a patient as nodule present and the proposed CAD indicates the absence of nodule, the detection test result would yield FN; and if the radiologist identifies a patient as nodule absent and the proposed CAD indicates the absence of nodule, the detection test result would yield true negative (TN). Using TP, FP, FN, and TN, various performance metrics such as classification accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F-measure, and G-mean are calculated.

8.2. Accuracy

Accuracy can be defined as the ratio between the sum of TP and TN to the total sum of attributes used. Accuracy relies mainly on the classification rate of the classifier.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{AN}{TOTAL} \quad \text{(I)}$$

8.3. Sensitivity

Sensitivity can be defined as the ratio between TP to the sum of actual positive. It defines how the nodule is correctly diagnosed.

$$\text{Sensitivity} = \frac{TP}{AP} \quad \text{(II)}$$

8.4. Specificity

Specificity is the ratio between TN to the sum of actual negative. It defines how well the absence of nodule is correctly diagnosed.

$$\text{Specificity} = \frac{TN}{AN} \quad \text{(III)}$$

8.5. Receiver operating characteristics

Receiver operating characteristics (ROC) is graphical representation between FP rate (FPR) and TP rate (TPR). FPR can also be defined as (1-specificity). In ideal situation, the sensitivity and specificity of diagnostic result will be 100% and this is called perfect classification.

8.6. Positive predictive value

The performance of proposed CAD and ground truth should predict correctly the prevalence of disease. Mathematically, positive predictive value (PPV) can be expressed as:

$$PPV = \frac{TP}{PP} \quad \text{(IV)}$$

8.7. Negative predictive value

The performance of proposed CAD and ground truth should predict correctly the absence of disease. Mathematically, negative predictive value (NPV) can be expressed as:

$$NPV = \frac{TN}{PN} \quad \text{(V)}$$

8.8. F-measure

F-measure can be defined as the weighted mean value of precision and recall.

$$\text{F-measure} = \frac{2^{*}(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad \text{(VI)}$$

8.9. G-mean

G-mean maintains a balance between the positive class and negative class classification accuracies. The classification

precision of a positive class is defined by sensitivity. In addition, specificity determines how well a negative class is classified. The value nearer to 100% represents the perfect classification accuracy.

$$G-mean=\sqrt{(sensitivity*specificity)} \quad \text{(VII)}$$

8.10. Mathew's correlation

In Mathew's correlation (MC), the actual and predicted condition takes the value between 0 and 1. The value of 1 corresponds to perfect correlation, whereas the value of 0.5 corresponds to random prediction.

$$\frac{(TP*TN)-(FP*FN)}{\left[(TP+FP)(TP+FN)(TN+FP)(TN+FN)\right]^{1/2}} \quad \text{(VIII)}$$
$$=\frac{(TP*TN)-(FP*FN)}{\left[AP*AN*PP*PN\right]^{1/2}}$$

9. Results and discussion

The images used for examining the proposed methodology were taken from the LIDC-IDRI, SPIE-AAPM Lung CT challenge, and hospitals. Nodule size between 3 mm and 30 mm were considered in this work. Most specifically, solid, part-solid, and non-solid nodules were chosen. In the LIDC-IDRI database, 71 exams are chosen. Out of 71 exams, 246 nodule case and 240 non-nodule cases were selected. In the SPIE-AAPM database, out of 70 only 35 exams were used. Among them, 28 nodule cases and 34 non-nodule cases are selected. About 36 CT images were acquired from hospitals. A total of 584 images were considered in this work, of which 292 belong to nodule and the rest belong to non-nodule cases. The input datasets are grouped into training set and testing set with 292 datasets each. All these databases were aimed to promote the development of the proposed CAD system.

The performance measure of each classifier for different inertia weights can be measure using accuracy, sensitivity, and specificity. Using confusion matrix, these measures can be calculated. Accuracy of each classifier can be obtained correctly by determining the ratio of the correctly classified and total number of samples. Sensitivity can be measured from the misclassified rate of nodule case to the total number of nodule case used. Specificity can be measured from the misclassified rate of non-nodule case to the total number of non-nodule case used. Tables 1-3 describe briefly the performance of classifiers based on confusion matrix. For solid features, the accuracy, sensitivity, specificity, PPV, NPV, F-measure, G-mean, and MC of various classifiers are noted.

10. Conclusions

In this work, we proposed an automated CAD system for classification of lung nodules using various classifiers from CT images. The classification of nodule and non-nodule patterns in CT is one of the most significant processes during the detection of lung nodule. The developed CAD systems consist of segmentation, feature extraction and classification. For segmentation, we used filters for effective extraction infected region. Later, we extracted features through features and fed into classifiers such as DS, RF, and BPNN. The experimentation was conducted on LIDC-IDRI dataset (Tables 3-5), and the results with BPNN outperformed those with DS and RF classifiers. The performance was measured using sensitivity, specificity, PPV, NPV, F-measure, and G-Mean.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization: Naveen HM
Investigation: Naveen HM
Methodology: Naveen HM
Formal analysis: Naveena C, Manjunath Aradhya VN
Writing – original draft: Naveen HM
Writing – review & editing: Naveen HM

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Akbari R, Ziarati K, 2011, A rank based particle swarm optimization algorithm with dynamic adaptation. *J Comput Appl Math.*, 235(8): 2694–2714.

   https://doi.org/10.1016/j.cam.2010.11.021

2. Aoyama M, Li Q, Katsuragawa S, *et al*., 2003, Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images.

Med Phys, 30(3): 387–394.

https://doi.org/10.1118/1.1543575

3. Armato SG, Giger ML, Moran CJ, *et al.*, 1999, Computerized detection of pulmonary nodules on CT scans. Radiographics, 19(5): 1303–1311.

https://doi.org/10.1148/radiographics.19.5.g99se181303

4. Armato SG 3rd, Hadjiiski L, Tourassi GD, *et al.*, 2015, Guest editorial: Lungx challenge for computerized lung nodule classification: Reflections and lessons learned. *J Med Imaging*, 2(2): 020103.

https://doi.org/10.1117/1.JMI.2.2.020103

5. Armato SG, McLennan G, Bidaut L, *et al.*, 2011, The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med Phys*, 38(2): 915–931.

https://doi.org/10.1118/1.3528204

6. Arumugam MS, Rao MV, 2006, On the performance of the particle swarm optimization algorithm with various inertia weight variants for computing optimal control of a class of hybrid systems. *Discrete Dyn Nat Soc.*, 2006: 079295.

https://doi.org/10.1155/DDNS/2006/79295

7. Choi WJ, Choi TS, 2014, Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Comput Methods Programs Biomed*, 113(1): 37–54.

https://doi.org/10.1016/j.cmpb.2013.08.01

8. Criminisi A, Shotton J, Bucciarelli S, 2009, Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI Workshop on Probabilistic Models for Medical Image Analysis. vol. 1. Rochester, Minnesota: MICCAI Society.

9. Cross GR, Jain AK, 1983, Markov random field Texture models. *IEEE Trans Pattern Anal Mach Intell*, 5(1): 25–39.

https://doi.org/10.1109/tpami.1983.4767341

10. Da Silva Sousa JR, Silva AC, de Paiva AC, *et al.*, 2010, Methodology for automatic detection of lung nodules in computerized tomography images. *Comput Methods Programs Biomed*, 98(1): 1–14.

https://doi.org/10.1016/j.cmpb.2009.07.006

11. Dai S, Lu K, Dong J, *et al.*, 2015, A novel approach of lung segmentation on chest CT images using graph cuts. *Neurocomputing*, 168: 799–807.

https://doi.org/10.1016/j.neucom.2015.05.044

12. Daneshmand F, Mehrshad N, Massinaei M, 2013, A new approach for froth image segmentation using fuzzy logic. In: First Iranian Conference on Pattern Recognition and Image Analysis (PRIA). New York City: IEEE.

13. Dawoud A, 2011, Lung segmentation in chest radiographs by fusing shape information in iterative thresholding. *IET Comput Vision*, 5(3): 185–190.

14. De Nunzio G, Tommasi E, Agrusti A, *et al.*, 2011, Automatic lung segmentation in CT images with accurate handling of the Hilar Region. J Digit Imaging, 24(1): 11–27.

https://doi.org/10.1007/s10278-009-9229-1

15. Deep G, Kaur L, Gupta S, 2013, Lung nodule segmentation in CT images using rotation invariant local binary pattern. *Int J Signal Image Process*, 4(1): 20.

16. Dehmeshki J, Amin H, Valdivieso M, *et al.*, 2008, Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach. *IEEE Trans Med Imaging*, 27(4): 467–480.

https://doi.org/10.1109/TMI.2007.907555

17. Dehmeshki J, Ye X, Lin X, *et al.*, 2007, Automated detection of lung nodules in CT images using shape-based genetic algorithm. *Computer Med Imaging Graph*, 31(6): 408–417.

https://doi.org/10.1016/j.compmedimag.2007.03.002

18. Delogu P, Cheran S, De Mitri I, *et al.*, 2005, Preprocessing methods for nodule detection in lung CT. In: International Congress Series. vol. 1281. Netherlands: Elsevier.

19. Dheepak G, Premkumar S, Ramachandran R, 2015, Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Method. *Int J Power Control Comput,* 7: 24–28.

20. Doi K, 2007, Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computer Med Imaging Graph*, 31(4): 198–211.

21. Dolejsi M, Kybic J, Polovincak M, *et al.*, 2009, The lung time: Annotated lung nodule dataset and nodule detection framework. In: SPIE Medical Imaging. Washington USA: International Society for Optics and Photonics.

22. Elizabeth D, Nehemiah H, Raj CR, *et al.*, 2012, Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image. *IET Image Processing*, 6(6): 697–705.

23. Enquobahrie AA, Reeves AP, Yankelevitz DF, *et al.*, 2007, Automated detection of small pulmonary nodules in whole lung CT scans. *Acad Radiol*, 14(5): 579–593.

24. Farag A, Abdelmunim H, Graham J, *et al.*, 2012, An AAM based detection approach of lung nodules from LDCT scans. In: 9th IEEE International Symposium on Biomedical Imaging (ISBI). New York City: IEEE.

25. Farag A, Ali A, Graham J, *et al.*, 2011, Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. New York City: IEEE.

26. Farag AA, Abdelmunim H, Graham J, *et al.*, 2011b, Variational approach for segmentation of lung nodules. In:

IEEE International Conference on Image Processing (ICIP). New York City: IEEE.

27. Gambhir S, Shepherd J, Shah B, *et al.*, 1998, Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol*, 16(6): 2113–2125.

28. Garro BA, Vazquez RA, 2015, Designing artificial neural networks using particle swarm optimization algorithms. *Comput Intell Neurosci*, 2015: 369298.

https://doi.org/10.1155/2015/369298

29. Golosio B, Masala GL, Piccioli A, 2009, A novel multithreshold method for nodule detection in lung CT. *Med Phys*, 36(8): 3607–3618.

https://doi.org/10.1118/1.3160107

30. Gomathi M, Thangaraj P, 2010a, A computer aided diagnosis system for detection of lung cancer nodules using extreme learning machine. *Int J Eng Sci Technol*, 2(10), 5770–5779.

31. Gomathi M, Thangaraj P, 2010b, A computer aided diagnosis system for lung cancer detection using support vector machine. *Am J Appl Sci*, 7(12): 1532.

https://doi.org/10.3844/ajassp.2010.1532.1538

32. Alves LG, Novo J, Campilho A, 2016, Hessian based approaches for 3d lung nodule segmentation. *Expert Syst Appl*, 61: 1–15.

https://doi.org/10.1016/j.eswa.2016.05.024

33. Gould MK, Donington J, Lynch WR, *et al.*, 2013, Evaluation of individuals with pulmonary nodules: When is it lung cancer. *Chest*, 143(5 Suppl): 93S–120S.

https://doi.org/10.1378/chest.12-2351

34. Grigorescu SE, Petkov N, Kruizinga P, 2002, Comparison of texture features based on Gabor filters. *IEEE Trans Image Process*, 11(10): 1160–1167.

https://doi.org/10.1109/TIP.2002.804262

35. Gu Y, Kumar V, Hall LO, *et al.*, 2013, Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recogn*, 46(3): 692–702.

36. Gudise VG, Venayagamoorthy VK, 2003, Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks. In: Swarm Intelligence Symposium. New York City: IEEE.

37. Hua P, Song, M, Sonka EA, *et al.*, 2011, Segmentation of pathological and diseased lung tissue in CT images using a graph-search algorithm. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. New York City: IEEE.

38. Jacobs C, Murphy K, Twellmann T, *et al.*, 2011, Computer-aided detection of solid and ground glass nodules in thoracic CT images using two independent cad systems. In: The Fourth International Workshop on Pulmonary Image Analysis.

39. Shen S, Bui AA, Cong J, *et al.*, 2015, An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Comput Biol Med*, 57: 139–149.

https://doi.org/10.1016/j.compbiomed.2014.12.008

40. Shen S, Sandham W, Granat M, *et al.*, 2005, MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE Trans Inform Technol Biomed*, 9(3): 459–467.

https://doi.org/10.1109/titb.2005.847500.

41. Shi, Y, Eberhart R, 1998, A modified particle swarm optimizer. In Evolutionary Computation Proceedings, 1998. In: IEEE World Congress on Computational Intelligence, The 1998 IEEE International Conference. New York City: IEEE.

42. Shih-Chung BL, Freedman MT, Lin JS, Mun SK, 1993, Automatic lung nodule detection using profile matching and back-propagation neural network techniques. *J Digit Imaging*, 6(1): 48–54.

43. Yuan J, 2013, Active contour driven by local divergence energies for ultrasound image segmentation. *IET Image Processing*, 7(3): 252–259.

https://doi.org/10.1049/iet-ipr.2012.0120

44. Zhou S, Cheng Y, Tamura S, 2014, Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images. *Biomed Signal Process Control*, 13: 62–67.

https://doi.org/10.1016/j.bspc.2014.03.010

45. Flohr TG, Schaller S, Stierstorfer K, *et al.*, 2005, Multi-detector row CT systems and image-reconstruction techniques. *Radiology*, 235(3): 756–773.

https://doi.org/10.1148/radiol.2353040037

46. Kawane K, Fukuyama H, Yoshida H, *et al.*, 2003, Impaired thymic development in mouse embryos deficient in apoptotic DNA degradation. *Nat Immunol*, 4(2): 138–144.

https://doi.org/10.1038/ni881

47. Shi J, Malik J, 2000, Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Machine Intell*, 22(8): 888–905.

https://doi.org/10.1109/34.868688

48. Bercoff J, Tanter M, Fink M, 2004, Supersonic shear imaging: A new technique for soft tissue elasticity mapping. *IEEE Trans Ultrason Ferroelectr Freq Control*, 51(4): 396–409.

https://doi.org/10.1109/tuffc.2004.1295425