

A Hybrid Approach for the Lung(s) Nodule Detection Using the Deformable Model and Distance Transform

Ayyaz Hussain¹, Mohammed Alawairdhi², Fayez Alazemi³, Sajid Ali Khan⁴ and Muhammad Ramzan^{2,*}

¹Department of Computer Science, Quaid e Azam University, Islamabad, 44000, Pakistan

²College of Computing and Informatics, Saudi Electronic University, Riyadh, 11673, Saudi Arabia

³Public Authority for Applied Education and Training, 6500, Kuwait

⁴Department of Software Engineering, Foundation University Islamabad, Islamabad, 44000, Pakistan

*Corresponding Author: Muhammad Ramzan. Email: m.ramzan@seu.edu.sa

Abstract: The Computer Aided Diagnosis (CAD) systems are gaining more recognition and being used as an aid by clinicians for detection and interpretation of diseases every passing day due to their increasing accuracy and reliability. The lung(s) nodule detection is a very crucial and difficult step for CAD systems. In this paper, a hybrid approach for the lung nodule detection using a deformable model and distance transform has been proposed. The proposed method has the ability to detect all major kinds of nodules such as the juxta-plueral, isolated, and the juxta-vesicular, along with the non-solid nodules automatically and intelligently. Results show an impressive 95.2% accuracy with 4.85 false positives per scan. One significant achievement of the proposed work is its ability to detect various sizes of nodules from 3 mm to 30 mm. The proposed technique has been tested on the publicly available Lung(s) Image Database Consortium (LIDC). The results clearly show the effectiveness of the proposed technique in early detection with impressive accuracy.

Keywords: Computer-aided diagnosis; segmentation; classification; lungs; pulmonary nodules; Juxta-Plueral; Juxta-Vesicular

1 Introduction

Research shows that lung cancer is one of the most prevalent types of cancer [1–2] and lung cancer is the most frequent cause of death among all the types of cancers [3–4]. Among the cancers, lung cancer has the second-lowest survival rate only topped by pancreatic cancer in the five-year relative survival data. The survival rates are less than 10% for both male and female [5]. Recent reports on lung cancer suggest that the continued existence rate of lung cancer in the past five years varies between 13% to 21% [6]. This rate shows an increase of up to 50% when lung nodules are diagnosed in early stages. As per statistics from 2008, lung cancer is a prevalent disease, which affects a very large population of people throughout the world. As per the last global estimates, annually a total of 1,200,000 new cases of lung infections are reported [7] and the numbers are on the rise. However, unfortunately despite all the latest technologies available to us, diagnoses are still regularly made late, affecting treatment results. Diagnosing lung cancer using low dose computed tomography is a big hope for changing this situation as such and would be more proactive and effective at arriving at conclusions about lung tumors.

Research shows two major reasons for the tendency of late diagnosis. First, using available methods, it is difficult to diagnose lung cancer at early stages due to insufficiency of symptoms. Second, poor prognostics also contribute to the problem. It is a hard challenge to diagnose early whether a pulmonary



nodule exists or not, and whether the nodule is malignant or not. During the diagnosis, radiologists need to analyze hundreds of Computed Tomography (CT) images relying solely on human judgment. This results in vulnerability and the enhanced probability of mistakes by the radiologists. The need of including automated intelligent support for the radiologists is spurring an increased interest in CAD systems [8–9].

The CAD-based diagnosis system supports findings made by radiologists based on quantitative analysis of the radiological image. The basic idea/steps used in the CAD schemes are first, processing of images for extraction and detection of the nodule candidates from the images. Second, image feature's quantitation is the candidates of abnormalities. Third, classification of the data between abnormal and normal features of lungs images (benign and malignant) and last, a quantitative assessment and recovery of pictures like those of obscure sores. In a lung's tumor detection, computed tomography (CT) imaging techniques are widely considered the most significant and most responsive method of detecting lung nodules. A figured tomography (CT) sweep is an imaging technique that uses X-beams to make pictures of a cross-area of the body. Automated nodule recognition plans have been shown to significantly increase indicative precision in radiological imaging [10–11]. Fig. 1 shows a description of the different structures of lungs [12].

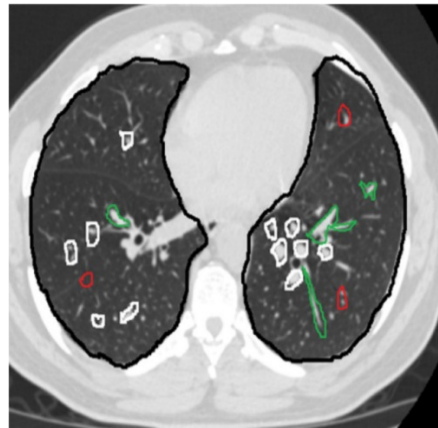


Figure 1: An image of a Lung CT scan where the white color shows Lung Nodules, and the green color represents the Pulmonary and the red color are the Human Airways

An important aspect, which needs to be considered is that a radiologist analysis is mainly based on the morphological structures under investigation, which can be checked in a 3D space. The examination of a CT is performed through the bi-dimensional picture so in essence there is a tradeoff between the radiologist's needs to be the perceptive and what is provided to him. This requires a remodeling of the tri-dimensional parts of the tissue under investigation. This is an intricate undertaking, which therefore leaves the door open for many mistakes. That's the major reason there is a great demand for computational frameworks, which helps with the task of discovery and analysis when it pertains to lung nodules [13]. A representation of a chest X-ray and CT image are shown in Fig. 2.

The paper is organized as follows: After the introduction, we present a brief and concise literature review outlining previous work and the current state of research in this domain. In Section 3, the proposed methodology is explained followed by a discussion of the experimental results achieved in Section 4. Section 5 is the conclusion and future work is described.

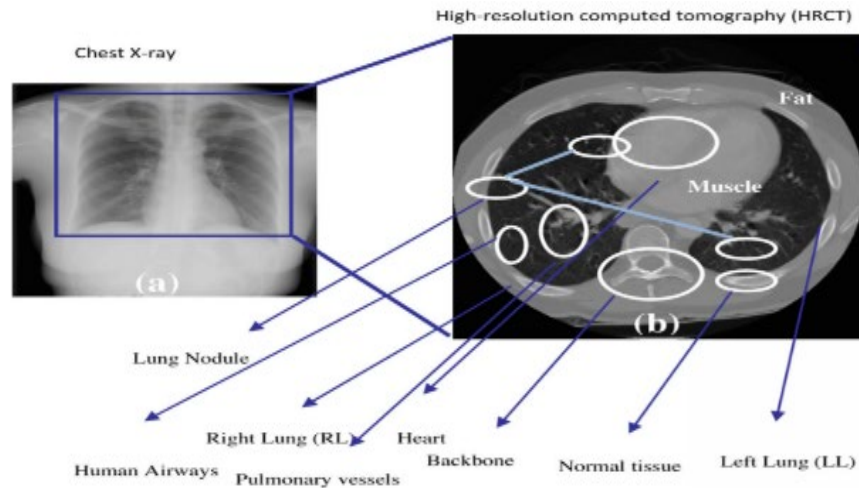


Figure 2: Representation of a Chest X-Ray image with Labelling

2 Literature Review

Literature studies show that Dehmeshki et al. [14] presented a CAD plan named as the hereditary calculation format coordinating for the programmed recognition of lung knobs. This was based on the premise of the geometric state of the voxels and with the worldwide conveyance of the knob power performs the calculation of the wellness capacity. Sousa et al. [8] presented a CAD plan to perform divisions at numerous stages. Every division stage was responsible for segments of the volume of CT image. Ye et al. [9] proposed a CAD sachem in which they utilized five elements containing force data, shape record and a 3D spatial area. Opfer et al. [15] demonstrated impressive results of the CAD-based identification with lung knob demonstrating 89% affectability in recognizing the knobs having sizes more noteworthy than 4mm and 60% affectability for having measurements under 4mm. Messay et al. [16] connected various dim level limits to the volumetric lung areas to distinguish knob hopefuls. Likewise, various format coordinating based routines have been worked on. Pei et al. [17] introduced a three-stage technique, utilizing a 2D multi-scale channel, separation of blob-molded knobs and on-knobs and lastly, extraction and classification of the shape elements. They performed testing on 30 exams and displayed an affectability of 100% and a false positive rate of 8.4 for each exam. The CAD proposed by Tan et al. [18] was based on the neural system. The hereditary calculation furthermore incorporates developments such as the utilization of another classifier of elements depicted as the highlight deselected neuro-advancing enlarging of the topologies. Two different classifiers were additionally utilized, i.e., the settled topology manufactured neural systems and the SVM. The model during testing showed the affectability calculation of (87.5%). Lee et al. [19] proposed a two staged method of classifiers called sporadic woods. In the first stage, the area of lung handles was chosen, and in the second stage, there was an effort to reduce the false positives. The results demonstrated by them showed a 100% accuracy for the bonafide positives and 1.4% for the false positives for every picture. Camarlinghi et al. [20] proposed the combination of various CAD frameworks for the upgraded performance to the lung knob ID. Their analysis was then contrasted with consequences of the individual frameworks by the method for the ROC curve. The results showed that the higher the quantity of the CAD frameworks utilized as a part of the identification, the higher the quantity of the genuine positives of 65, and the lower the quantity of the false positives of 139 and over the LIDC base of 69 pictures containing 114 lung knobs. Chama et al. [21] presented a system that uses mean-movement taken after by the methods in view of geometric properties, for example, the Region of interest (ROI), made from the symmetric centered guide of two typical topics. They performed experimentation on 429 pictures, (133 ordinary and 296 strange) from the LIDC-IDRI from the Interstitial Lung Disease (ILD) databases. The proposed technique accomplished sensitivities and specificities of 97% and 99%, (ordinary pictures) and 83% and 99%, (unusual pictures), separately.

Choi et al. [22] introduced a hereditary programming-based component change and grouping for the programmed identification of the pneumatic knobs on processed tomography pictures. This CAD approach worked in three fundamental steps. In the initial step, the lung division was performed utilizing the thresholding and 3D part marking. In the second step, the ideal thresholding and manage based pruning for the location of the knob applicants are utilized and in the last step, a GP-based classifier is used to sort the knobs and non-knob elements.

Recently with the development of the convolution neural network (CNN), more and more researchers [23–28] used the CNN for detection of lung nodules. Juang et al. [29] used an Automatic Multi-Thresholding for the tumor classification. Similarly, Xu et al. [30] also proposed a novel technique for the medical image segmentation using the self-adaptive PCNN model.

The major contributions in this paper are the following: (1) The proposed technique does not focus on a single type of nodule. Instead, it detects all types of lung nodules including isolated nodules, juxta-plural nodules, and juxta-vesicular nodules. (2) It can detect the juxta-vesicular nodules of a large size, which were not detected earlier in any technique due to their complex structure. The Juxta-vesicular nodules are complex to work with, because they are attached to other structures such as vessels and the bronchi tree.

The proposed scheme is based on a four-step approach, which works as follows: Initially a lung segmentation is performed using a linear interpolation and lung parenchymal identification technique. In the second stage, we detect the Region of Interest (RoI) using multiple thresholds for the ROI, contour correction and a seed point choice. After detection, in the third stage, the process of the nodule detection is carried out using both the deformable model for detection (for Juxta-plural nodules and isolated nodules) and the Distance Transform (for juxta-vesicular nodules). In the final stage, the false-positive reduction is achieved using Fuzzy rule-based pruning.

This brief overview of literature shows that despite impressive developments in this field, there are still significant shortcomings hampering the effectiveness of the CAD based systems in a true sense. The limitations we are focusing on in this paper include a comprehensive mechanism that can work with various kinds of nodules simultaneously. Another area of interest for us on which no previous work has focused on so far deals with the large size juxtavesicular nodules. In the Section 3, we explain the structure of proposed methodology.

3 The Proposed Methodology

The lung(s) segmentation is the process to detect the lungs part from the whole CT Scan image of the lungs. There are many different organs available in the complete CT scan lung image, therefore it is important to segment the lungs for the detection of the nodules available only in the lungs. The lung(s) segmentation, nodule detection is a very important and crucial step to extract the nodules. There are four types of lung(s) nodules that are required to detect. The proposed methodology consists of four steps to perform the whole process.

- The lung parenchymal and linear interpolation techniques are used to performed the Lung Segmentation.
- The Region of Interest (RoI) has been detected using a multiple threshold.
- The nodule detection is carried out using the deformable model (for Juxta-pleural nodules and isolated nodules) and the Distance Transform (for juxta-vesicular nodules).
- The false-positive reduction is achieved using the Fuzzy rule-based pruning.

Graphical representation of our proposed methodology is given below in Fig. 3.

In the following subsections, we explain the complete process of each stage separately in detail.

3.1 Linear Interpolation

The performance of any CAD system depends greatly on the examination based on the volume of the Computer Tomography pictures. For the first step, we transform all the voxels into a 3D coordinate system network with uniform 3D spatial determination. This helps overcome the risk of blunder due to the

anisotropic representations of lattices. This approach used in our work, despite its simplicity, is a strategy that is broadly used for informal recreation. Keep in mind that the spatial determination along the hub heading in the CT examination is not the same as the spatial determination inside every cut. That is why we have used a straight addition along the hub bearing as shown in the image.

3.2 Lung Parenchymal Volume Identification

In any CAD scheme the lung segmentation is necessary, because to detect the nodules from the input image is a difficult task. In our work, the input image is segmented into several slices to minimize the computation complexity and the FPs. We calculate the initial volume of the interest (VOI) for separation of the lung(s) portions. The objective is to better identify the juxta-vesicular nodules, the isolated nodules and the juxta-pleural nodule. The segmentation makes it easier to find different types of nodules from the image. The most used technique for the lung segmentation in the literature is the threshold-based region filling. The limitation of this technique is that it erroneously removes the important regions including the juxta-pleural nodules, because these nodules are attached with the pleura. The simple thresholding filling technique or algorithm is not able to detect the juxta-pleural nodules separately. To overcome this limitation, the proposed technique works on following two steps:

- The Inclusion Process

A 3D Region Growing algorithm is used for the lung parenchyma identification.

- The Connectivity Analysis

To incorporate the interior structures with the high-power esteem (knobs and vessels) and the pleura layer neighboring inward lung volume, a widening procedure is connected, which is called the Connectivity analysis.

3.2.1 The Inclusion Process

When we examine the structure of the lungs, we find that the bronchial tree and air lies in the interior lung volume. In the CT Images, these appear as a low power voxel encompassed by the high force voxels, which are connected to the pleura surface. We use this information to divide the interior lung volume using a 3D RG (Region Growing) calculation. Our technique begins from a seed point, which is the incremental appending, so it combines adjacent focuses, which have the comparable properties to the seed. This process keeps iterating until there are no more focuses fulfilling the predefined criteria. The most challenging task in the RG (Region Growing) calculation is to pick a better incorporation process with a proper seed point. For this purpose, we have utilized the Simple Bottom Threshold (SBT). In this technique, if the power of a nonexclusive voxel is lower than the edge esteem, this voxel is added to the area. The termination point is consequently chosen by the system proposed by Ridler et al. [31]. This procedure depends on the CT voxel dim quality circulation, which is comprised of two parts: One has air, bronchial tree, trachea, and lung parenchyma. The other is comprised of muscles, bones, vascular tree, and fat. The ideal edge is set between these two locales. The RG calculation utilizes the two seed focuses, and determines the criteria of the seed point as described by Cascio et al. [32] shown as follows:

1. The start point on the central voxel (P0) is set first.
2. Then left of a voxel, a shift is performed.
3. Start the RG for such voxel, if the rule of inclusion is satisfied, then go to the Step 2.
4. If the magnitude order of the selected volume and expected lung volume are the same at the end of the RG process, then end the trial phase, otherwise go to Step 2.

A point of the lung volume is chosen from the histogram of the CTs dark tones and evaluating the territory under the second top from the left. This part relates to the voxels having a place with the lung parenchymal and bronchi. To select the required volume of pulmonary, all the above steps are repeated. Below Fig. 3 displays an example of the inclusion process.

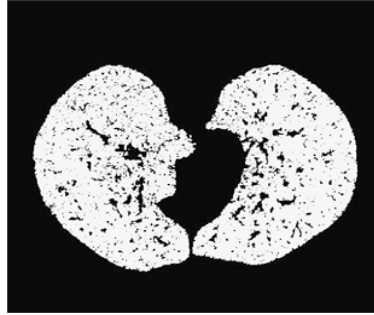


Figure 3: Image after the inclusion process

3.2.2 The Connectivity Analysis

After the inclusion process, we receive the segmented image. However, in this image the internal nodule, vessels and air walls are not included. In order to include these nodules, a dilation process called the connectivity analysis is applied. For an accurate application, the connectivity analysis is applied. After applying the connectivity analysis, the juxta-pleural nodules are also included into the segmented image. Fig. 4 represents the proposed methodology diagram.

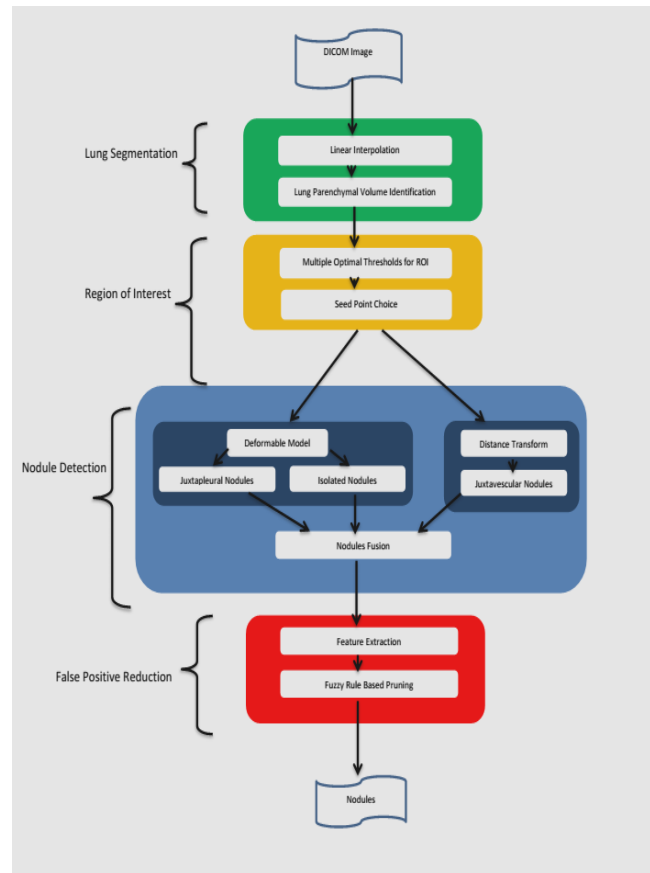


Figure 4: The proposed methodology diagram

3.3 The Multiple Optimal threshold for the ROI Extraction

The process of selecting multiple optimal thresholds is shown in Fig. 5. As we know, there are different intensities of nodules and varying levels of vessel attachments. In order to extract the Region of Interest

(ROI) in such a situation, we use multiple thresholds. We have used the optimal threshold to detect the ROI. The Thoracic CT consists of two main groups of pixels, a high-intensity pixel located in the body (body pixels) and a low-intensity pixel, which are in the lungs and surrounded air, which are called the non-body pixels. The larger intensity difference between these two group of pixels ensures the thresholding is the best mechanism to separate these groups. The method proposed by Choi et al. [22] used five static threshold values, which is not an optimal threshold. In our technique we use the method of the optimal thresholding defined by Hu et al. [33]. This technique iteratively figures the estimation of a limit so that the two gatherings of the pixels are all around isolated.

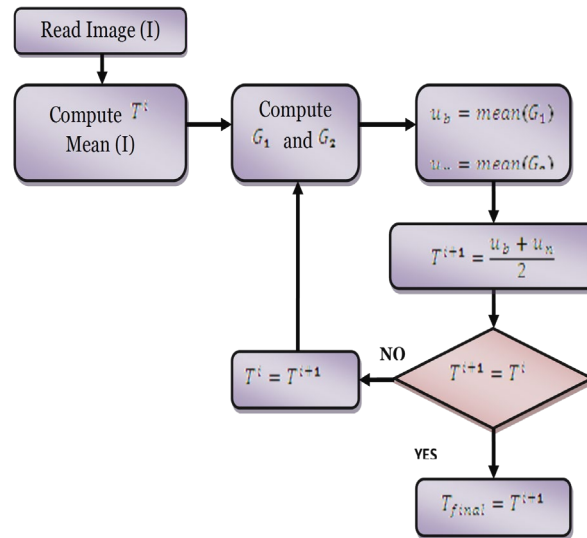


Figure 5: The multiple optimal threshold selection process

Let T^w be the edge esteem at step w and μ_a, μ_b is the normal power estimation of the body pixels, (i.e., with the force higher than T^w), and separately the non-body pixels (power lower than T^w). The limit for step $w + 1$ is:

$$T^{w+1} = \frac{\mu_a + \mu_b}{2} \tag{1}$$

This technique is rehashed until the union, i.e., and until step e , where $T^e = T^{e-1}$. The introductory limit of T^0 is set to 128, which is the middle dim level. At the point when joining is finished, the picture is thresholded at quality T^e . The body pixels are set to 0 and the non-body pixels are set to 1.

3.4 The Seed Points Choice

The seeds chosen work as the information for the step is called the “Deformable Model”. The knobs have a more prominent force as for the pneumonic parenchyma and they can be effortlessly found by searching for the neighborhood maxima in the volume of interest. A voxel-level operation is performed as follows:

$$O(a,b,c) = (M(a,b,c) - N(a,b,c)) \cdot I(a,b,c) \tag{2}$$

where $M(a,b,c)$ is the matrix after the connectivity analysis process, $N(a,b,c)$ is the matrix we get after the RG process, $I(a,b,c)$ is the input image, and $O(a,b,c)$ is the resultant matrix. Fig. 6 represents the result of Eq. (1). After getting the matrix $O(a,b,c)$, we apply a hit the highest point detector algorithm on it to discover the confined maxima and select the seed points, which will be used in the next step. The process of the Seed Point Choice is shown in Fig. 6.

3.5 The Nodule Detection

In this stage, we detect all the various types of nodule candidates from other structures. To detect these nodules, we used two techniques. The first one is the deformable model to detect the isolated nodules and the juxta-pleural nodules and the second one is the distance transform to detect the juxtavascular nodules. Detailed description of these methodologies is given below.



Figure 6: The resultant image after the seed point choice

3.5.1 The Deformable Model

This model is used to detect the isolated and juxtapleural nodules from the lung parenchyma. This model is also called the Mass Spring Model. The basic aim of the deformable model is to identify the shape of the object. We have used it to identify the nodule's shape among other structures. The Deformable model uses a priori knowledge by definition and parameterization. The Deformable model discovers the balance between the inner strengths (portraying the model shape) and the outer powers (depicting the picture data). In the material science-based displaying worldview, the shape relates to the connected (outside) powers on the deformable model so that the model converges into the shape of an object, while the inside powers keep the model smooth amid disfigurement. The deformable model is adjusted considering the vitality term until the vitality is insignificant.

The Design of the Deformable Model

We need to locate a suitable vitality capacity from which we get an answer for the knob form. This model was introduced by Cascio et al. [32]. In the first stages of the 3D deformable model, we took a circular cross-section of mass. The model begins from an augmented position where the mass cross-section holds on to the potential knobs. The seed focuses, which were picked in the past section are taken as the focal point of every circle network. We marked the circle with the N value of the mass focus that make the dynamic model and with the t number of the mass purposes of a solitary cut that makes a connection of $L = t.u$, where u is the total number of slices. Each mass in the sphere $N(j)$ ($j = 1 \dots t$) is connected with two other masses $N(j - 1)$, $N(j + 1)$, which belongs to the same slice and with two slices $N(j - n)$ and $N(j + n)$, which belongs to the previous slice. We have written it in a more generalized way as:

$$E_{functional}^{tot} = \sum_{i=1}^N [E_{internal}(i) + E_{external}(i)] \quad (3)$$

As shown in Eq. (2), every single element in the proposed model plays an important role with external and internal energies to the functional energy. Now we will describe internal and external forces in detail.

Internal Energy

The internal energy $E_{internal}(i)$ of the deformable model is achieved by the calculation and addition of all following energies.

Elastic Energy

In order to get the shape of the segmented object, the first step is the Elastic energy. The vertex i is connected to the z -plane and xy -plane of its four neighbors. Those four points are contributing to the elastic energy.

$$E_{elas}(i) = \alpha[E_{spr,i-1}(i) + E_{spr,i+1}(i) + E_{spr,i-n}(i) + E_{spr,i+n}(i)] \quad (4)$$

where $E_{spr j}(i)$ represents the elastic energy of the model, which is attached with the two masses i and j .

$$E_{spr j}(i) = \frac{1}{2}k\Delta x^2 \quad (5)$$

The two-mass points distance is represented by Δx and k , which denotes the spring constant, as a parameter.

Bending Energy

The second step is to apply the bending energy on the object. The model must adjust to the diverse bends that the object might show. In this way we apply the bending energy. Here the bending energy is linked with the general vertex I , which is defined as follows:

$$E_{bend} = \beta\|P_{i-1} - 2P_i + P_{i+}\|2 \quad (6)$$

where P_i indicates the position of the ith vertex.

Attraction Energy

The vertices belong to the same slice, and the geometric centre of those positions are taken as a linking point for the calculation to separate the vertices normal (\bar{d}) and standard deviation (σ_d). If the ith vertex has the distance of d_i from the center, then it is represented as:

$$d_i > \bar{d} + \sigma_d \quad (7)$$

Then the attraction energy is linked with the ith vertex and given as:

$$E_{attr}(i) = \gamma \frac{d_i}{\bar{d}} \quad (8)$$

With respect to the average distance as the vertex is from the center caused the larger amount of energy contribution. This contribution makes the points on the model surface almost spherical, which are most likely nodules.

External Energy

Internal energy evaluates the contour points as well as the contraction and bending but it does not provide any pulling force with which we can separate our desired object from the rest of the image [32]. Therefore, we apply external forces for this task. The external energy $E_{external}$ is the combination of the gradient and potential energies. Now we describe the contribution of each energy in our work one by one respectively.

Gradient Energy

Gradient energy has a high gradient evolution towards the location. A very small amount of energy will be associated with the ith vertex if it is located on the edge point, which is represented by Eq. (9) [32]:

$$E_{grad}(i) = -\delta\|\nabla I(P_i)\| \quad (9)$$

where $I(P_i)$ is the intensity of the ith vertex at the P_i position.

Potential Energy

Every point of the model is linked with potential energy. If the point is with high intensity, then it means more energy is linked with it. If a pixel moves from a low-intensity value to a large intensity value like the pleura surface, then it must pay the additional energy function such as:

$$E_{pot}(i) = I(P_i) \quad (10)$$

Energy Functional

After calculating the internal and external forces we must sum up all the energies to get the total energy function, which is represented as follows:

$$E_{functional} = \alpha E_{elas}(i) + \beta E_{bend}(i) + \gamma E_{attr}(i) + \delta E_{grad}(i) + \epsilon E_{pot}(i) \quad (11)$$

Dimensional parameters like ϵ , β , α , γ and δ are empirically reduced. Various sizes of the artificial objects are introduced in the image for identification of these parameters. At the start of the methodology, we mentioned that our aim is to detect the isolated and juxta-pleural nodules, and both types of nodules were detected successfully.

Now for the detection of the juxta vesicular nodules, we use the distance transform technique. A detailed description of this technique is given.

3.5.2 The Distance Transform

After applying the deformable model, we get the isolated, juxta-pleural and some small juxta-vesicular nodules. The deformable model excludes the larger juxta-vesicular nodules due to their complex structure. To include those larger juxta-vesicular nodules, we apply the distance transform using the 3D region growing algorithm. We used the seed points, which were identified earlier. As a matter of the first importance, the structures are regrouped by utilizing the 3D district developing calculation. This helps in joining the structures with one another. Initially, the resultant structures are not partitioned into the genuine structures (knobs, vessels and bronchi).

The shape of the knobs is the most important element, which is utilized to distinguish it from the different structures. It provides us with essential data about the shape and we use this data for analysis. The difficulty arises from distinguishing the round shapes (more probable knobs) from the barrel-shaped shapes (more probable vessels). The distance change is figured [8] in each of the structures in all the voxels of the edges. As we traverse to the inward voxel, it is increased until there are one or two voxels left that are distinguished from the edge. The structures with the expansive grouping of the knobs have a high estimation of the separation change. This information about the lung knob is used to separate the bigger juxta-vesicular knobs from alternate structures, for example, the vessel and bronchi and so on.

3.5.3 The Nodules Fusion

After detection of all the types of nodule candidates at this stage, we merge all those types of nodules before performing the rule-based pruning. We take union of all the detected nodules to avoid duplication and to make the rule-based pruning much easier.

3.6 The False-Positive Reduction Using the Fuzzy Rule-Based Pruning

There is probability of non-nodule elements present in the earlier stage detection module. These non-nodule elements are unnecessary and affect the accuracy of the system. Rule based pruning has been performed and four rules are defined to remove these unwanted non-nodules. Nodule candidates are characterized by their volume (v), area (a), elongation (e), diameter (d) and circularity (c). Maximum and minimum threshold has been defined and is used for the distinguishing nodules from non-nodules elements.

For the symbolic representation of minimum and maximum thresholds of volume, area and diameter $T_v^{(min)}$, $T_a^{(min)}$, $T_d^{(min)}$, $T_v^{(max)}$, $T_a^{(max)}$ and $T_d^{(max)}$ values are used for nodule candidates respectively. This threshold is used non-nodule candidates' separation. Objects falling below or above threshold are considered as non-nodule candidates and objects values lying within the threshold are categorized as nodule candidates.

Pruning Rules:

Rule R₁:

Diameter (l) $< T_d^{(min)}$

Rule R₂:

Volume (l) $> T_v^{(max)}$ or Overlapped (l, VS) $> T_0$ or ($Elongation (l) > T_e$ and $Volume (l) > T_v^{(min)}$)

Rule R₃:

Diameter (l) $> T_d^{(max)}$

Rule R₄:

Circularity (l) $> T_c$ and $Area(l) > T_a^{(min)}$ and $Area(l) < T_a^{(max)}$.

All these four rules can help us in separating the non-nodules candidates from nodule candidate. The fuzzy rule-based pruning algorithm is presented here:

```

1. function Pruning (NFs)                → Remove vessels and noise (Nodule Fusion(NF))
2. for each NF ∈ NFs do
3.   NC ← ∅                               → Nodule Candidates
4.   VS ← ∅                               → Vessels
5.   if Diameter(l) < Td(min) then
6.     continue                          → Noise
7.   end if
8.   if Volume (l) > Tv(max) or Overlapped (l, VS) > T0
   or (Elongation (l) > Te and Volume (l) > Tv(min) ) then
9.     VS ← VS ∪ {l}                     → Vessels
10.    continue
11.  end if
12.  if Diameter (l) > Td(max) then
13.    continue                          → Noise
14.  end if
15.  if Circularity (l) > Tc
   and Area(l) > Ta(min) and Area(l) < Ta(max) then
16.    NC ← NC ∪ {l}                       → Nodule Candidates
17.  end if
18. end for
19. return NC
20. end function

```

4 The Results/Comparisons with Other Works

Nodules are mainly categorized into three types namely juxta-vascular, juxta-plueral and isolated-nodules. Techniques developed by [22] have addressed the problem of detection of all types of nodules. Their technique was unable to detect some of the large and non-solid nodules. However, there is deterioration in accuracy of juxta-vascular nodules detection. It is only because of the larger size and complicated structure of these types of nodules. Our proposed technique has been successful in categorizing larger juxta-vesicular type nodules along with some non-solid nodules. In order to detect the larger juxta-vesicular nodules, which were not detected through earlier techniques, we used the distance transform technique. As shown in Tab. 1, three different distribution of 30–70%, 50–50% and 70–30% are used for training and testing purposes.

Table 1: The results of the proposed nodule detection technique

Training-Testing	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
30-70	92.45	82.58	99.23	0.9616
50-50	96.25	96.42	99.64	0.9938
70-30	97.23	96.82	99.72	0.9926

Accuracy, specificity, and sensitivity measurements are used to check the system performance with training to testing ratio. Tab. 1 illustrates that the accuracy and sensitivity gradually increased with an increase in the training sample. However, the specificity rate, which shows the relationship between the

false positive and the true negative are almost consistent with an increase in the training samples. The best results are obtained with 70–30% training to testing ratio.

It is important to compare the results of the proposed method in the same domain with existing methods. It is a challenging task due to the change in the evaluation parameters like the results metrics, nodule size and dataset. For fair comparison we have considered the methods using the same experimental protocols as shown in Tab. 2. The number of the false-positive rate is the most important evaluation parameter, which greatly affects the accuracy of any good diagnosis system. It is also the key requirement of any diagnosis system to reduce the false-positive rate. Tab. 2 shows the clear picture that the results of the proposed method are better than the existing method with respect to the average false positive rate and sensitivity. Suzuki et al. [34] achieved 80.3% sensitivity with a 16.1% false-positive rate. The methods reported in [16] and [22] used the same size nodule as ours, however, their sensitivity rate is less than our proposed method.

Table 2: The comparative analysis of the proposed nodule detection technique

CAD system	Nodule size used (mm)	Average false positives per case	Sensitivity
Suzuki et al. [34]	8–20	16.1	80.3
Opfer et al. [15]	≥ 4	4	74
Messay et al. [16]	3–30	3	82.66
Choi et al. [22]	3–30	5.45	94.1
Proposed CAD	3–30	4.85	95.2

Fig. 7 shows the proposed method performance in terms of the scatter plot. It shows the classification of correct and incorrect nodules. The figure depicts that the proposed method has reduced the false positive rate and increased the true positive rate.

The results clearly show the performance improvement using our proposed technique for the false-positive detection against most of the standard techniques. At the same time, we can see that the efficiency of the proposed technique is much better than most of the already existing methodologies as shown in Fig. 8. Fig. 8 shows the illustration of the ROC curve, the AUC and the results of the current classifier. The results in Fig. 8 also show the different features for each group, which are described in terms of false positives and true positives. Our proposed method improved the performance in terms of the true positives, sensitivity, accuracy and specificity.

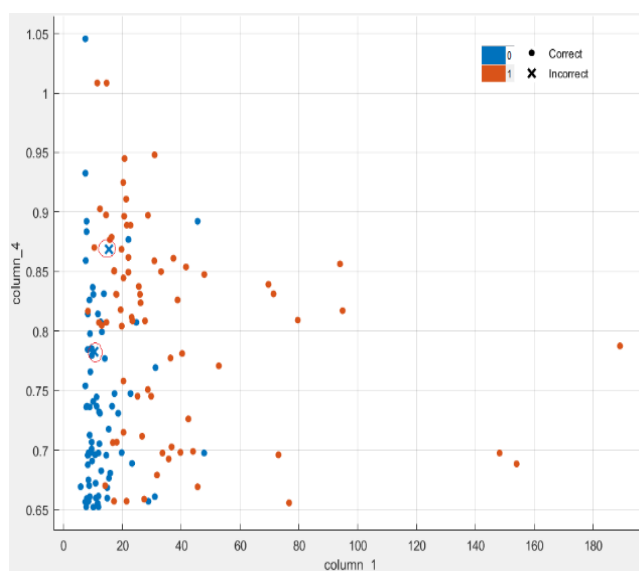


Figure 7: The scatter plot of the proposed method

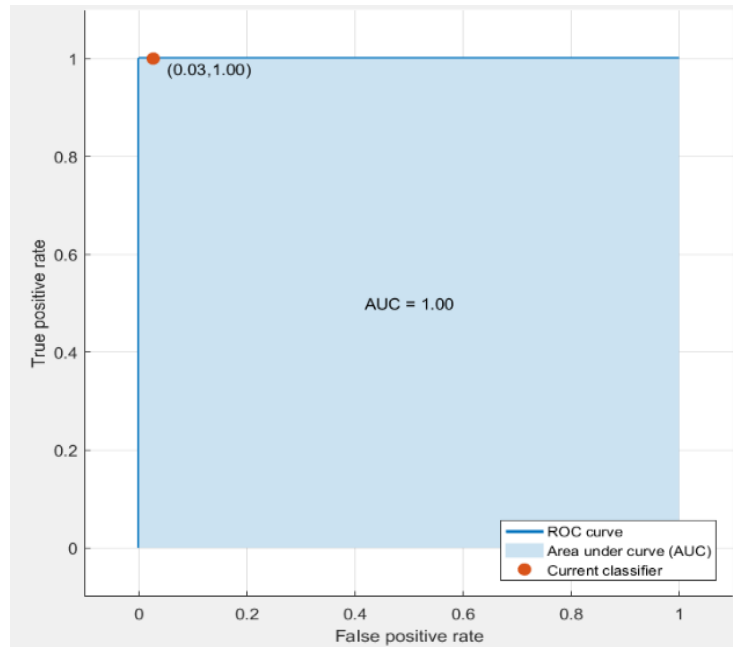


Figure 8: The false-positive ate using the proposed method

5 Conclusion and Future Work

There is a significant rise in the use of CAD systems to help radiologists detect lung nodules and increase the survival rate of lung cancer patients. These systems are very helpful for the radiologists and patient's survival rate has been drastically increased. However, there are few limitations of these types of systems when applied using CT-Scan images. A novel framework for CAD Systems is developed and is presented in this research paper. Lungs are segmented using Linear Interpolation and Lung Parenchymal Volume Identification methodology. We have used multiple optimal thresholding method for region of interest extraction. Candidate nodules are detected by applying deformable model and distance transform. Proposed technique performs better than the existing methods and is capable of detecting all types of nodules. System detects malignant and benign lesions. However, categorization of benign and malignant is left. In our experiments we have used LIDC datasets and results are generated. It has been observed that there is significant improvement in sensitivity, i.e., 95.2% and deterioration in false positive rate, i.e., 4.85%.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. A. Khan, K. Kenza, M. Nazir and M. Usman, "Proficient lungs nodule detection and classification using machine learning techniques," *Journal of Intelligent & Fuzzy Systems*, vol. 28, no. 2, pp. 905–917, 2015.
- [2] S. A. Khan, S. Hussain, S. Yang and K. Iqbal, "Effective and Reliable Framework for Lung Nodules Detection from CT Scan Images," *Scientific Reports*, vol. 9, no. 1, pp. 1–14, 2019.
- [3] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu and M. J. Thun, "Cancer statistics, 2009," *CA: A Cancer Journal for Clinicians*, vol. 59, no. 4, pp. 225-249, 2009.
- [4] K. W. Jung, Y. J. Won, S. Park, H. J. Kong, J. Sung *et al.*, "Cancer statistics in Korea: Incidence, mortality and survival in 2005," *Journal of Korean Medical Science*, vol. 24, no. 6, pp. 995–1003, 2009.
- [5] V. Beral and R. Peto, *UK Cancer Survival Statistics*, British Medical Journal Publishing Group, 2010.

- [6] J. A. Barta, C. A. Powell and J. P. Wisnivesky, "Global epidemiology of lung cancer," *Annals of Global Health*, vol. 85, no. 1, pp. 8, 2019.
- [7] T. Ferkol and D. Schraufnagel, "The global burden of respiratory disease," *Annals of the American Thoracic Society*, vol. 11, no. 3, pp. 404–406, 2014.
- [8] J. R. F. da Silva Sousa, A. C. Silva, A. C. de Paiva and R. A. Nunes, "Methodology for automatic detection of lung nodules in computerized tomography images," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 1, pp. 1–14, 2010.
- [9] X. Ye, G. Beddoe and G. Slabaugh, "Graph cut-based automatic segmentation of lung nodules using shape, intensity, and spatial features," in *Workshop on Pulmonary Image Analysis, MICCAI*, London, UK, 2009.
- [10] H.P. Chan, E. Charles, P. Metz, K. L. Lam, Y. Wu *et al.*, "Improvement in radiologists' detection of clustered microcalcifications on mammograms," *Arbor*, vol. 1001, no. 1, pp. 48109–50326, 1990.
- [11] D. J. Getty, R. M. Pickett, C. J. D'Orsi and J. A. Swets, "18 Enhanced interpretation of diagnostic images," *Judgment and Decision Making: An Interdisciplinary Reader*, pp. 281, 2000.
- [12] Q. Abbas, "Segmentation of differential structures on computed tomography images for diagnosis lung-related diseases," *Biomedical Signal Processing and Control*, vol. 33, no. 1, pp. 325–334, 2017.
- [13] C. Villaseñor, N. Arana-Daniel, A. Y. Alanis and C. Lopez-Franco, "Differential evolution and covariance ellipsoid for non-rigid transformation tracking of internal organs," in *2018 Int. Joint Conf. on Neural Networks (IJCNN)*, Rio, Brazil, 2018, pp. 1–7.
- [14] J. Dehmeshki, X. Ye, M. V. Casique and X. Lin, "A hybrid approach for automated detection of lung nodules in CT images," in *3rd IEEE Int. Sym. on Biomedical Imaging: Nano to Macro*, USA, pp. 506–509, 2006.
- [15] R. Opfer and R. Wiemker, "Performance analysis for computer-aided lung nodule detection on LIDC data," in *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*, vol. 6515, 65151C, 2007.
- [16] 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.
- [17] X. Pei, H. Guo and J. Dai, "Computerized detection of lung nodules in CT images by use of multiscale filters and geometrical constraint region growing," in *2010 4th Int. Conf. on Bioinformatics and Biomedical Engineering*, Chengdu, China, 2010, pp. 1–4.
- [18] M. Tan, R. Deklerck, B. Jansen, M. Bister and J. Cornelis, "A novel computer-aided lung nodule detection system for CT images," *Medical Physics*, vol. 38, no. 10, pp. 5630–5645, 2011.
- [19] S. L. A. Lee, A. Z. Kouzani and E. J. Hu, "Automated identification of lung nodules," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, Cairns, Queensland, Australia, 2008, pp. 497–502.
- [20] N. Camarlinghi, I. Gori, A. Retico, R. Bellotti, P. Bosco *et al.*, "Combination of computer-aided detection algorithms for automatic lung nodule identification," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no.3, pp. 455–464, 2012.
- [21] C. K. Chama, S. Mukhopadhyay, P. K. Biswas, A. K. Dhara, M. K. Madaiah *et al.*, "Automated lung field segmentation in CT images using mean shift clustering and geometrical features," *Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670, 867032, 2013.
- [22] 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, no. 1, pp. 57–78, 2012.
- [23] L. Gong, S. Jiang, Z. Yang, G. Zhang and L. Wang, "Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 11, pp. 1969–1979, 2019.
- [24] G. Kasinathan, S. Jayakumar, A. H. Gandomi, M. Ramachandran, S. J. Fong *et al.*, "Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier," *Expert Systems with Applications*, vol. 134, no. 1, pp. 112–119, 2019.
- [25] J. Tan, Y. Huo, Z. Liang and L. Li, "Expert knowledge-infused deep learning for automatic lung nodule detection," *Journal of X-ray Science and Technology*, vol. 27, no. 1, pp. 17–35, 2019.
- [26] Q. Wang, F. Shen, L. Shen, J. Huang and W. Sheng, "Lung nodule detection in CT images using a raw patch-based convolutional neural network," *Journal of Digital Imaging*, vol. 32, no. 6, pp. 971–979, 2019.

- [27] 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, no. 1, pp. 109–119, 2019.
- [28] W. Zuo, F. Zhou, Z. Li and L. Wang, "Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection," *Ieee Access*, vol. 7, no. 1, pp. 32510–32521, 2019.
- [29] L. H. Juang and M.N. Wu, "Tumor classification using automatic multi-thresholding," *Intelligent Automation & Soft Computing*, vol. 24, no. 2, pp. 1–9, 2017.
- [30] X. Xu, T. Liang, G. Wang, M. Wang and X. Wang, "Self-adaptive PCNN based on the ACO algorithm and its application on medical image segmentation," *Intelligent Automation & Soft Computing*, vol. 23, no. 2, pp. 303–310, 2017.
- [31] T. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 8, pp. 630–632, 1978.
- [32] D. Cascio, R. Magro, F. Fauci, M. Iacomi and G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models," *Computers in Biology and Medicine*, vol. 42, no. 11, pp. 1098–1109, 2012.
- [33] S. Hu, E. A. Hoffman and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 490–498, 2001.
- [34] K. Suzuki, "A review of computer-aided diagnosis in thoracic and colonic imaging," *Quantitative Imaging In Medicine and Surgery*, vol. 2, no. 3, pp. 163, 2012.