Applying Ant Colony Optimization algorithms and variants for lung nodule detection

K. Veerakumar
Assistant Professor, Department of ECE, RVS College of Engineering and Technology, Dindigul
veeramtech5@gmail.com

C G Ravichandran
Principal, Excel Engineering College, Komarapalayam
cgravichandran@gmail.com

K.R. Aravind Britto
Associate Professor, Department of ECE, RVS College of Engineering and Technology, Dindigul
krbritto1975@gmail.com

Abstract
Ant Colony Optimization (ACO) algorithms are widely used in medical imaging, especially for image edge detection and image segmentation. In this paper, fundamentally we use ACO algorithm for lung nodule detection and compare the performance against three other algorithms namely Otsu algorithm, Watershed algorithm, Global region based segmentation. In addition, we suggest a novel approach which involves variations of ACO namely Refined ACO, where number of iterations are extended based on performance parameters, Logical ACO, where outputs of every refined iterations are XORed to be fed for subsequent iterations and Variant ACO, where the outputs of Otsu algorithm becomes the input of refined ACO.

Keywords Lung cancer, lung nodules, lung CT scan, edge detection, machine learning.

1 Introduction
Lung cancer is one of the most common types of cancers in the world. The impact of lung cancer differs based on the stages of its occurrence. Lung cancer is generally detected based on the presence of lung nodules. Lung nodules are very common and can be seen in 1 out of 500 scans. They are small masses of tissue in the lung. Not that all lung nodules cause cancer. A nodule which does not cause cancer is called benign and cancerous nodules are referred as malignant. In medical imaging, computer aided detection of lung nodules serves as a hot area of research.
Lung nodule detection initially involves edge detection from the lung images. There are many methods to detect edges from the images. Basically there are two categories namely, gradient, which detects the edges from the first derivative maximum and minimum values and laplacian, which uses the second derivative zero crossing of the image. There are three basic operators which forms the basis for edge detection. They are Sobel operator, which detects edges using 3x3 convolution matrices, Robert cross operator, which detects edges using 2x2 convolution matrices and Prewitt operator, which is similar to Sobel operator and is used to detect vertical and horizontal edges in image. Applying machine learning algorithms for detection and segmentation of irregular shaped lung nodules will remain a remarkable milestone in CT scan image analysis research.

In this paper we discuss about the application of ACO algorithms for pulmonary nodule detection and suggest some variations in the Ant Colony Optimization algorithm to arrive at better results. Other algorithms namely, Otsu algorithm, Watershed algorithm and Global region based segmentation algorithm are compared with the performance of ACO. In addition, we propose three levels of improvements to ant colony optimization algorithm. First level is extending the number of iterations based on its performance criteria thereby refining the output of normal ACO. The second is that the refined outputs are X-ORed to make the output more logical. The third level alteration involves giving the output of another algorithm, say Otsu, as input to ACO algorithm.

2 Related work

Computer aided detection systems are widely used to detect and diagnose numerous diseases and is a prime area of research in medical imaging. Most common anatomical regions covered under CAD are lung, chest, breast and brain. Summers RM says that CT or MRI scan images are popularly used as input for such detection systems [25]. Doi Et. al., clarifies the popular and significant applications include locating clustered micro-calcifications in mammograms, and pulmonary nodule detection for lung cancer in [6]. Most commonly knowledge based approaches provide better results in CAD applications. Commonly used classifiers are: Rule-based, k-NN, ANN, Decision trees, Naïve Bayes, LDA and SVM. Out of these, Korfiatis Et. al., [14] presents rule-based classifiers which are simple to implement, and can be integrated with other classifiers to produce better results. However, selection of disease identification threshold has to be done manually.

K nearest neighborhood approach given by Chen and Zrimec T Et. al., [3, 31] provides consistent classification results. But this method has the problem of identifying the correct ‘k’ value which shows improved accuracy with larger datasets. Artificial neural networks proposed by Pun CM Et. al.,[23] are yet another remarkable milestone in medical imaging applications. The completeness and consistency of performance provided by ANN based applications are due to the ability of system to learn itself completely from the input-output combinations independent of domain specific issues. The reason for less popularity of ANN is that this required heavy dose of training to guarantee error-free and consistent learning, which is not always possible. Alternatively, decision tree based learning attempted to grab the attention of ANN researchers in medical imaging is said by Kauczor HU [12].
Unlike ANNs, decision tree learning has low computational complexity and less training. However, over-fitting was the potential drawback in decision trees. Neither ANN nor Decision trees were able to solve the problem of CAD of diseases. Then came the optimal approach called Naïve Bayes classification, where the problems of both ANN and Decision trees were ruled out. Yet another scheme is given by Cheng HD Et. al., for classification came into existence, much simpler than Naïve Bayes, known as LDA [4]. In addition, this approach was analytically simple and computationally less expensive. LDA demands selection of appropriate feature sets of input images for better classification results. The most improved optimal solution for classification is support vector machines. Though this approach has very high algorithmic complexity, the results of SVM training and classification was close to human inferences.

CAD of pulmonary nodules works with two phases: pulmonary nodule detection and nodule pattern and nodule shape identification. Nodules are generally irregularly round and opaque. They may be solid, non-solid or partly-solid and exist with less than 3 cm diameters [10]. Nodular patterns are similar to nodules but vary from 2-10 mm in diameter and are generally widely spread over the lung regions. In addition, micronodules are less than 3mm in diameter. A disease affected lung region shall contain one or many or combination of nodules and nodular patterns. Therefore, computer aided detection of nodules and nodular patterns includes analysis of shape based features in CT images. Studies made by Boroczky L Et. al., reveal that size, volume, area, diameter, 2D and 3D dimensions, circularity, solidarity, thickness, top-hat filtering, mean curvature, shape index, gaussian curvature, sphericity, surface smoothness, shape irregularity, roundness, center of mass, compactness, inertia matrix and surface curvature are the useful and effective features for pulmonary lung nodule detection and nodule pattern analysis [1,3,9,13,21,27,28,29,30]. In addition, histogram based approaches and gradience information are also used in detection of small cavities in radiograph images. Several studies by Shen R Et. al., [24] include combination of shape based and texture based methods for pulmonary lung nodule detection.

Ant colony algorithms are widely applied in wireless networks, especially in routing and load balancing [15]. Very little research focus is put on applying ACO algorithms to medical imaging [19]. These include application of ACO to lymph node classification [5], brain tumour segmentation [22,18], hippocampus segmentation [8], prostate cancer classification [21], diabetes diagnosis [7], retinal vessel detection [11], ovarian cancer detection [26], and heart ventricle segmentation [16]. Significantly visible research has been performed over applying ACO for micro-calcification detection of mammograms by Lochanambal, K[17]. A remarkable milestone in CAD of lung nodules is reported by Bram van Et. al.,[2]. This work combines the output of various algorithms for automated pulmonary nodule detection of CT images and obtains better improvement in performance. In this context, this paper experiments ant colony based approaches for automatic lung nodule detection and combines the output of various ACO and non-ACO algorithms to obtain remarkable performance improvements.

3 Existing approaches for lung nodule detection
In this section, we present four existing algorithms for edge detection. Since lung nodule detection fundamentally involves edge detection, we use edge detection algorithms as the primary step of lung nodule detection. The detailed working of the algorithms is presented below.

3.1 Otsu algorithm

Otsu uses gray scale images for its image processing steps. Hence we take a grayscale lung CT images for further processing. In this method it searches for the pixels with all possible threshold values and finds the spread of pixels in each threshold range. It involves finding the pixels that fall under foreground and background. The edge is detected when the sum of foreground and background spread is the maximum. The mean weight and variance are calculated. Then within class variance is calculated whose value is used to detect the edge (refer algorithm 1). Figure 1a & 1b shows the input-output of Otsu edge detection.

Algorithm 1: Otsu algorithm

begin
  for each pixel of image
    find weight of pixel for both foreground and background
    find mean of weights for both foreground and background
    find variance of the pixels
  end for
  calculate within class variance for all the pixels
  maximum value of within class variance gives the edge pixels.
End

Fig.1 a Input Lung CT image

Fig.1 b Otsu output image

3.2 Watershed algorithm

In grey scale images, different grey levels indicate the edges. Watershed algorithm basically sees the image as topographic relief. The basic idea behind this is construction of dams. The catchment areas refer to the object we are trying to segment; here the catchment areas are the lung nodules. As the water level increases, dams are constructed to protect
ourselves. When the water level reaches the highest peak construction stops. Same way, we start from the watershed pixels and grow iteratively. When the edge detected reaches the maximum level, the process stops and gives the required edges. The detailed procedure is given in algorithm 2. Figure 2b represents the edge detection of Watershed algorithm for figure 2a.

**Algorithm 2: Watershed algorithm**

begin
convert image into binary image
compute the euclidean distance transform of the binary image
identify the watershed regions from the image
for each watershed region
   label each pixel with an value
   label with value 0 indicates it does not belong to unique threshold region
   watershed pixels are pixels with value 0
end for
create a sobel horizontal filter
filter the input image with the sobel filter to get the regions.
end

3.3 Global region based segmentation algorithm

The lung fields are segmented in CT image using a region growing algorithm. The algorithm is based on the selection of pixel, the pixel can be selected either by giving (x,y) co-ordinate or clicking a pixel from the CT image. After selecting the pixel, the regions associated with this pixel based on connectivity and gray scale difference were formed by using the region mean. Through this method the given CT image were segmented and lung nodule edges are detected. The procedure of global region based segmentation is given in
algorithm. For figure 3a, the input image, the edge detection by global region based segmentation is shown in figure 3b.

**Algorithm 3: Global region based segmentation algorithm**

begin
Label seed points according their initial groping
Put neighbours of seed points (the initial T) in the Sequentially Sorted List (SSL).
While the SSL is not empty
  Remove first point y from SSL.
  Test the neighbours of this point:
  if all neighbours of y which are already labelled (other than with the boundary label) have
    The same label- set y to this label.
  else
    Update running mean of corresponding region add neighbours of y which are
    neither already set nor already in the SSL according to their value of \( \delta \)
  end if
end while
end

![Fig.3a Input Lung CT image](image1)
![Global Region-Based Segmentation](image2)

**3.4 Ant Colony Optimization algorithm**

The basic idea behind this algorithm is the movement of ants. All ants follow the same path with the help of pheromone which will be left by the preceding ants. The succeeding ants make use of this pheromone to find its path. Therefore each ant incrementally constructs a solution to the problem. Likewise, every ant constructs the edge of input image which is done iteratively to obtain the edges of lung nodules. The detailed procedure is given in
Algorithm 4: Ant colony optimization algorithm

begin
Initialize the base attractiveness, $\tau$, and visibility, $\eta$, for each edge
for $i < \text{IterationMax}$ do:
    for each ant do:
        choose probabilistically (based on previous equation) the next state to move into;
        add that move to the tabu list for each ant
        repeat until each ant completed a solution
    end
for each ant that completed a solution do:
    update attractiveness $\tau$ for each edge that the ant traversed
end
if (local best solution better than global solution)
    save local best solution as global solution
end
end

Fig. 4a Input Lung CT image  Fig. 4b Ant Colony Optimization output image

4 Proposed novel ACO approach for lung nodule detection

We propose three different levels of alterations of normal ACO algorithm. Initially we extend the number of iterations for lung nodule detection, then the refined outputs are XORed to improve its output value and finally tried to give the output of another edge detection algorithm, we have used Otsu algorithm, to be fed as input to the ACO algorithm. A detailed explanations and outputs are given in the following section.

4.1 Refined Ant Colony Optimization algorithm
From the results of normal Ant Colony Optimization algorithm, we notice some improper edges are detected. To overcome this, the idea of giving the output of previous iteration as the input to the next iteration helps. Since we are refining the iterations we get a better output than the normal Ant Colony Optimization algorithm. This is continued until the differences of subsequent iterations do not alter much. The pseudocode is presented in algorithm 5. Figure 5b shows the improved edge detection results for the image in figure 5a.

**Algorithm 5: Refined Ant Colony Optimization algorithm**

begin
  Initialize threshold for performance
  do
    apply Ant Colony Optimization algorithm (given in Algorithm 4)
    get the output of previous iteration as the input to the next iteration
    compare two subsequent outputs
    while (threshold is not equal to difference value)
    output of last iteration is the output of refined Ant Colony Optimization algorithm
  end

4.2 Logical Ant Colony Optimization algorithm

From the refined Ant Colony Optimization algorithm's output and normal Ant Colony Optimization algorithm's output, we notice that it detects noises along with the nodules. Hence a logical operation is applied to get even better detection of lung nodules. We get the final iteration output of refined ACO and the previous iteration output of refined ACO algorithm and then apply XOR to it to get the logical ACO output. This shows further reduction of noises in the output image. The detailed pseudocode is given in algorithm 6.

**Algorithm 6: Logical Ant Colony Optimization algorithm**

begin
get outputs from the refined Ant Colony Optimization algorithm
for each pair of refined ACO output
apply XOR to get input
repeat until all outputs are XORed
end for
end

Fig. 6a,b Logical ACO input (Refined ACO final and pre-final output images)  
Fig. 6c Logical ACO output image

4.3 Variant Ant Colony Optimization algorithm

To further reduce the unwanted noise in the output image, we chose to give the output of already run edge detection algorithm as the input to the normal ACO which becomes the Variant Ant Colony Optimization algorithm. The procedure includes Otsu algorithm and is explained in algorithm 7. The improved edge detection is shown in figure 7b.

Algorithm 7: Variant Ant Colony Optimization algorithm

begin
  feed the otsu algorithm output as the input for Ant Colony Optimization algorithm
  apply Ant Colony Optimization algorithm to the image
  the output image is the output of logical Ant Colony Optimization algorithm
end

Fig. 7a Input of Variant ACO (It is the output of Otsu algorithm)  
Fig. 7b Variant ACO output image
5. Results and evaluation

All the existing and proposed algorithms are evaluated based on recall ratio for a set of 33 lung CT images. The evaluation is done by getting the no. of nodules actually present in the original image and in the output image of each algorithm. A Recall value is calculated for evaluation purpose. It is the ratio between number of nodules in the output and the original image.

\[
\text{Recall} = \frac{\text{No. of nodules retrieved}}{\text{No. of nodules actually present}}
\]

If recall =1, it is accurate, if recall <1, it finds only a few and if recall >1, unnecessary noise is also detected as a nodule. Finally a graph is plotted to get the graphical representation for easier verification. Fig. 8 shows the evaluation of all 7 algorithms based on recall ratio.

The X-axis indicates the image ID and Y-axis indicate the recall value which is a ratio. Fig. 8 a shows the evaluation graph of Otsu algorithm. The recall ratio value is majorly distributed near 1, which indicates the performance of Otsu is good. Fig.8 b shows the evaluation graph of Watershed algorithm. The recall value is spread over arrange because of some over segmentation problems. Fig.8c shows the evaluation graph of global region based segmentation algorithm. The distribution of the recall ratio indicates it still involves improvisation. It can be achieved by careful selection of seed points.

Fig. 8d shows the evaluation graph of Ant Colony Optimization algorithm. Here most of the recall ratio is situated below 0.5 which indicates much more work has to be done for enhancing the performance. The improvisation work carried out is evaluated further. Fig. 8e shows the evaluation graph of Refined Ant Colony Optimization algorithm. It still has to be improved but gives better performance than the normal Ant Colony Optimization algorithm. It helps in the reduction of noise detection.

Fig. 8f shows the evaluation graph of logical Ant Colony Optimization algorithm. It visibly shows better performance but still there are wrongly detected nodules. It indicates distribution of recall value above 1. Fig. 8g shows the evaluation graph of Variant Ant Colony Optimization algorithm. Though there is distribution of the recall ratio over a range, it shows better performance above all the proposed approaches.
Fig. 8 a Evaluation of otsu algorithm

Fig. 8 b Evaluation of Watershed algorithm

Fig. 8 c Evaluation of Global Region based Segmentation algorithm

Fig. 8 d Evaluation of Traditional ACO algorithm

Fig. 8 e Evaluation of Refined ACO algorithm

Fig. 8 f Evaluation of Logical ACO algorithm
Specificity and sensitivity are also used for the evaluation purpose. Specificity measures the proportion of negatives while sensitivity measures the proportion of positives correctly identified. Sensitivity = TP/(TP+FN) and Specificity = TN/(TN+FP), where TP is number of true positives, FP is number of false positives, TN is number of true negatives and FN is number of false negatives. Fig. 9 shows the evaluation of all 7 algorithms based on specificity and sensitivity.

The figures show that Otsu, watershed and global region based segmentation algorithms are able to detect nodules positively from the sensitivity graphs and specificity graphs show a little degradation in the detection of nodules because they include portions which are not actually nodules. The variations in ACO algorithms are better in detection of nodules than the normal ACO algorithm. The average sensitivity and specificity values of all the images are taken and plotted for all seven algorithms. Fig. 10a shows the sensitivity graph while 10b shows the specificity graph of average values of all 33 images. It can be used to easily figure out the overall performance.
Fig 9c  Sensitivity of Watershed algorithm

Fig 9d  Specificity of Watershed algorithm

Fig 9e  Sensitivity of Global region based Segmentation algorithm

Fig 9f  Specificity of Global region based Segmentation algorithm

Fig 9g  Sensitivity of Ant Colony Optimization algorithm

Fig 9h  Specificity of Ant Colony Optimization algorithm
Fig 9i  Sensitivity of Refined ACO algorithm

Fig 9j  Specificity of Refined ACO algorithm

Fig 9k  Sensitivity of Logical ACO algorithm

Fig 9l  Specificity of Logical ACO algorithm

Fig 9m  Sensitivity of Variant ACO algorithm

Fig 9n  Specificity of Variant ACO algorithm
Accuracy and precision are also used to evaluate the process. Accuracy measures the degree of closeness of measurement of a quantity to the actual value of the quantity. Precision is the degree to which repeated measurements under unchanged condition shows the same result. It is represented as,

\[
\text{Accuracy} = \frac{TP+TN}{TP + FN + TN + FP}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

Fig. 11 gives the graphical representation of accuracy of all the seven algorithms with their average accuracy value of the 33 images. It depicts that the existing algorithms show a significant accuracy values while in the proposed algorithms, variant ACO shows higher accuracy than other ACOs. Fig. 12 shows the graphical representation of average precision of the algorithms. It significantly shows that, variant ACO is better in detection of lung nodules than all other algorithms. Logical ACO is not good in detection because we use the output of refined ACO as the input. This reduces precision value further.
Fig. 11 Histogram of average accuracy value of all seven algorithms

Fig. 12 Histogram of average accuracy value of all seven algorithms
5 Conclusion

This paper discussed a bold and novel attempt in applying ant colony optimization algorithm to pulmonary nodule detection, which is a breakthrough in ACO research in medical imaging. We also suggested various improvements to ACO in three levels: iteration based, logical and hybrid. Further, variant ACO shall be extended to include the output of other edge detection algorithms. In addition, SVM based training shall be attempted at the input to obtain better and clearer edge detection and classification of modified ACO.

References

2. Bram van Ginneken a,b,*, Samuel G. Armato III c, Bartjan de Hoop d, Saskia van Amelsvoort-van de Vorst d, Thomas Duindam a, Meindert Niemeijer a, Keelin Murphy a, Arnold Schilham a, Alessandra Retico e, Maria Evelin Fantacci e,f, Niccolò Camarlinghi e,f, Francesco Bagagli e,f, Ilaria Gori e,g, Takeshi Hara h, Hiroshi Fujita h, Gianfranco Gargano i,j, Roberto Bellotti i,j, Sabina Tangaro j, Lourdes Bolaños k,l, Francesco De Carlo j, Piergiorgio Cerello k, Sorin Cristian Cheran k, Ernesto Lopez Torres l, Mathias Prokop, Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study, Medical Image Analysis 14 (2010) 707–722


