



Lung nodule classification using artificial crawlers, directional texture and support vector machine



Bruno Rodrigues Froz^{a,*}, Antonio Oseas de Carvalho Filho^a, Aristófanés Corrêa Silva^a, Anselmo Cardoso de Paiva^a, Rodolfo Acatauassú Nunes^b, Marcelo Gattass^c

^a Federal University of Maranhão, Av. dos Portugueses, SN, Campus do Bacanga, Bacanga, São Luís, MA 65085-580, Brazil

^b State University of Rio de Janeiro, São Francisco de Xavier, 524, Maracanã, Rio de Janeiro, RJ 20550-900, Brazil

^c Pontifical Catholic University of Rio de Janeiro, R. São Vicente, 225, Gávea, Rio de Janeiro, RJ 22453-900, Brazil

ARTICLE INFO

Article history:

Received 12 January 2016

Revised 5 September 2016

Accepted 17 October 2016

Available online 17 October 2016

Keywords:

Artificial life

Artificial crawlers

Rose diagram

Lung nodule classification

ABSTRACT

Lung cancer is the major cause of death among patients with cancer throughout the world. The main symptom that indicate the lung cancer is the presence of lung nodules. This work proposes a methodology to classify lung nodule and non-nodule using texture features. The state-of-art of the presented work are the adaption of the Artificial Crawlers and Rose Diagram techniques for representing patterns over 3D images. Several information are extracted based on the texture behavior of these methods, allowing the correct classification of lung nodules candidates using Support Vector. Objective: This work proposes a methodology to classify lung nodule candidates and non-nodule candidates based on computed tomography (CT) images. Methodology: The Lung Image Database Consortium (LIDC-IDRI) image database is employed for our tests. Three techniques are employed to extract texture measurements. The first technique is artificial crawlers (ACs), an artificial life algorithm. The second technique uses the rose diagram (RD) to extract directional measurements. The third technique is a hybrid model that combines texture measurements from artificial crawlers and the rose diagram. The support vector machine (SVM) classifier with a radial basis kernel is employed. Results: In the testing stage, we used 833 scans from the LIDC-IDRI database. For the application of the methodology, we decided to divide the whole database into two groups, training and testing. We used partitions of training and testing of 20/80%, 40/60%, 60/40% and 80/20%. The division was repeated 5 times at random. We reached a mean accuracy (mACC) of 94.30%, a mean sensitivity (mSEN) of 91.86%, a mean specificity (mSPC) of 94.78%, a coefficient of accuracy variance (CAV) of 1.61% and a mean area under the receiver operating characteristic (mROC) curves of 0.922. Conclusion: Lung cancer has the highest mortality rate and one of the smallest survival rates after diagnosis. An early diagnosis increases the survival chance of patients. The proposed methodology is a useful tool for specialists in the detection of nodules. We believe we contribute for the expert system field because 1) the adaption of the Artificial Crawlers and Rose Diagram methods as 3D texture descriptors is innovative and contains great potential; 2) we adapted and developed measurements from the 3D texture descriptors; and 3) the simplicity and discriminative power of the methodology can be extended to applications based on images with other contexts.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Cancer is an uncontrolled growth of cells in a specific region of the body. Lung cancer has the highest mortality rate after diagnosis because of its late detection and treatment (Wood et al., 2012).

The high mortality rate is associated with its late diagnosis, which normally occurs at advanced stages of the disease.

The presence of lung nodules suggests cancer and indicates the stage of the disease. The detection of nodules is essential to patient prognosis and an early diagnosis increases the chance of survival. Additionally, increased information increases the precision of diagnosis. Interest in the development and use of digital image processing techniques applied to CT, with the goal of enhancing diagnosis precision, has increased recently. The combination of these techniques has led to the development of computer-aided

* Corresponding author. Fax: +559832728841.

E-mail addresses: bfroz63@gmail.com (B.R. Froz), antoniooseas@gmail.com (A.O. de Carvalho Filho), ari@dee.ufma.br (A.C. Silva), paiva@deinf.ufma.br (A.C. de Paiva), rodolfoacatauassu@yahoo.com.br (R. Acatauassú Nunes), mgattass@tecgraf.puc-rio.br (M. Gattass).

detection (CAD)/computer-aided diagnostic systems (Polakowski et al., 1997).

A CAD system typically consists of four stages: 1) image acquisition; 2) segmentation of nodule candidates; 3) extraction of features from the candidates; and 4) reduction of false positives (classification into nodules and non-nodules). All stages are fundamental to the success of a CAD system. In this work, we examine the feature extraction and classification stages. In these stages, we obtain the necessary information to reduce the number of false results.

In CAD methodologies, the feature extraction stage depends on: 1) geometry (for example, how round the candidate is); and 2) texture (grey level distribution). In this work, we only use texture features.

The goal of this work is to analyze lung nodules, classifying the candidates into nodules or non-nodules. The features used for classification are extracted in three ways. The first technique uses an artificial life model, called artificial crawlers. The curves extracted from it use the candidate CT scans as environments. The second technique uses gradient vectors to extract texture features from the candidates based on a circular histogram called the rose diagram. The third technique is a hybrid model that combines the texture features extracted from the other two methods. The support vector machine classifier is used in this methodology.

In Section 2, we describe related work. In Section 3, the three proposed methods (ACs, RD and hybrid) are described. In Section 4, we outline the proposed methodology for classification of nodule candidates using the three methods and the support vector machine (SVM) classifier. In Section 5, we present and discuss the results. In Section 6, we report the conclusions of this work.

2. Related works

As previously explained, a lung nodule detection system is commonly divided into several steps. Each of these steps has its own challenges to overcome. In nodule segmentation, one of the biggest difficulties is in isolating the nodule when connected with a vessel (juxtavascular) or with pleura (juxtaleural), without loss of texture. With such losses, the whole system can be prejudiced, leading to wrong conclusions about the nodules.

After segmentation, wrong candidates to be nodules may be generated, known as false positives. One big challenge is to reduce the false positives without losing actual nodules. Determining which measurements of texture and/or form should be used in combination with a computational intelligence system is the main difficulty of any system.

Over the years, several researchers have been making efforts to develop methodologies to overcome these challenges. We briefly present some works that aim to contribute solutions to these problems.

Mousa and Khan (2002) used SVM to classify lung nodules and non-nodules in CT exams using circular features and the mean variance of the detected edges. They used 50 nodule and non-nodule candidates for training and 16 images for testing. An 87.5% increase in sensitivity was achieved by the SVM classifier. The cases of training and test was manually chosen for both nodule and non-nodule candidates.

Jing, Bin, and Lianfang (2010) presented a methodology that classifies lung nodules in CT scans based on shape, geometry, grey level and texture features using SVM. They used a rule-based approach to exclude blood vessels from the images before acquiring the region of interest (ROI). The extracted shape feature is the invariant moment (ellipticity and major and minor axis of the ROI). The geometric features are: area, circularity, slenderness and degree of rectangularity. The grey-level features are the mean and variance of the grey levels. The texture features are:

energy, entropy, contrast and adverse moment. The highest accuracy obtained was 85.44%. This work used 50 slice CT images, including 50 lung nodules candidates and 204 non-nodules candidates. The authors explain that the small sized candidates and the irregular shapes of the nodules caused a few omissions and misclassification.

Lee, Kouzani, and Hu (2010b) used a Random Forest based classification aided by clustering for detection of lung nodule. This method used a hybrid random forest algorithm to classify the lung nodules and non-nodules. The images were obtained from 32 patient scans in the LIDC-IRDI. The authors achieved 98.33% sensitivity and 97.11% specificity. The clustering techniques chosen was k-means and expectation maximization algorithms, which requires parameterization. The number of clusters used was 2 in the classification aided by clustering method.

Namin, Moghaddam, Jafari, Esmail-Zadeh, and Gity (2010) present a methodology with two steps, segmentation and classification of lung nodules in malign and benign using fuzzy k-nearest neighbour (FKNN). FKNN is used in both steps. The extracted features are textures based on the intensities of the voxels (Hounsfield units) and the geometry. An 88% accuracy and a mean of 10.3 false positives per exam were reported. The author also presented that the nodules with small size and/or irregular shapes were the main shortcomings of the segmentation method.

Tartar, Kilic, and Akan (2013) suggest an approach that classifies lung nodules in CT scans using hybrid features. Four techniques are explored: principal component analysis (PCA) and minimum redundancy maximum relevance (mrMR), statistical features extraction, geometrical features extraction and an hybrid method. They evaluated statistical values from PCA in two dimensions and mrMR with geometrical features. The results were 90.7% accuracy, 89.6% sensitivity and 87.5% specificity. The methodology adopts a dataset of 2D CT slices evaluated by radiologists, divided in 95 nodules and 75 non-nodules from 63 patients. The use of two dimensional images for shape-based features extraction may raises the sensitivity, but prejudice the specificity of the third and fourth method because of the limited analysis of the whole nodule property on the original 3D CT.

Zhang et al. (2013) proposed an approach to classify lung nodules in four categories: well-circumscribed, vascularized, juxtaleural and pleural tail. The method describes both the lung nodule and the surrounding context information to perform the characterization in two distinct ways: 1) superpixel labelling, which labels the pixels as foreground or background, and 2) context curve calculation. The image database resource was the early lung cancer action program (ELCAP), which contains 50 low-dose scans. This method resulted in 82.5% accuracy. Their database contained 379 unduplicated lung nodules with locations indicated by annotations. The nodules images was cropped with a 31 x 31 pixels window for extracting the context curves. As superpixel labelling results in texture characteristics, there is no guarantee that the nodule image sliced comprises all fundamental features of the nodule texture.

The methodology developed by Choi and Choi (2014) uses Hessian matrices to calculate two angular histograms. The angular histograms of the normal surface are used as features to eliminate structures linked to lung parenchyma and to classify the remainder as nodules or non-nodules. Using 84 exams from the LIDC, they achieved 97.4% accuracy, 97.2% sensitivity and 97.7% specificity. The method multi-scale dot enhancement filtering is applied into 3D shape of the object, but lacks of pre-processing techniques. It was used 84 exams with a total of 148 isolated, juxtaleural and juxtavascular nodules.

Carvalho Filho et al. (2014), proposed the classification of lung nodules and non-nodules using taxonomic diversity indices. The computation of the indices was based on phylogenetic trees, which were applied to the characterization of the nodule candidate. The

SVM classifier and the radial-basis function (RBF) kernel were used. The LIDC (833 scans) resulted in 97.55% mean accuracy, 85.91% mean sensitivity and 97.7% mean specificity.

In the work developed by Pu et al. (2008a) the authors use adaptive borders algorithms to delimit the area of the juxtaleural nodules without selecting lung external tissue together with the region of interest. The methodology is divided into two steps. The first one is the image pre-processing. The second one corrects the defects caused by the exclusion of juxtaleural nodules. The experiments used 20 datasets and obtained an accuracy of 90%. This work uses an Adaptive Border Marching method to segments the lung region, reducing the over-segmentation problem. However, this algorithm, although adaptive, requires some parameters, assigned arbitrarily in this work.

The methodology proposed by Shen, Bui, Cong, and Hsu (2015) presents a lung nodule segmentation method, focused on juxtaleural nodules. The authors separated the methodology into three steps. First, the image pre-processing generates one initial mask with an adaptive threshold. Secondly, the inflexion point detection (horizontally and vertically) is realized using chain-code algorithm, to find inflexion points. Lastly, correction of the lung edge is done using SVM. The methodology achieved an accuracy of 92.6%. The chain-code technique to detect inflexion points is highly sensitive to noises, which is the main challenge of this work. Depending on the scans quality, this technique strongly depends on the adaptive threshold method and SVM performance to not raise the over-segmentation and under-segmentation ratios.

Tan, Deklerck, Jansen, Bister, and Cornelis (2011) presents a CAD system to detect lung nodules in CT images using nodule and vessel enhancement filters and a computed divergence feature for cluster the nodules. To classify, invariant features were extracted, defined on a gauge coordinates system, that separates nodules from non-nodules structures, using SVM and Artificial Neural Network. This methodology reached an sensitivity of 87.5%. 235 cases of LIDC database were selected for evaluation.

The work of Pu, Zheng, Leader, Wang, and Gur (2008b) proposes a method to achieve lung nodule detection based on geometric analysis of the signed distance field in CT images. The methodology achieved an maximum sensitivity of 95.1%. The methodology was evaluated with 52 low-dose screening CT exams.

The related works have been described with their highlights and shortcomings. In this work, we proposed a methodology that overcomes those inconveniences. The majority of those works, with the exception of Carvalho Filho et al. (2014), Shen et al. (2015) and Tan et al. (2011), utilized few exams for their study cases. This work takes the advantage of use a relevant number of exams, inspired by Carvalho Filho et al. (2014) work. Shen et al. (2015) chain-code approach is very sensitive to noises, differently from the proposed method of Artificial Crawlers, which is low sensitive to them, due to its evolutionary properties. The work of Mousa and Khan (2002) uses manually segmented nodules and non-nodules candidates, in contrast of the proposed methodology. The nodules candidates are segmented using markers by the radiologists and the non-nodules are obtained from Carvalho Filho et al. (2014) methodology, since the non-nodule separation in this work is automatically generated. Jing et al. (2010) and Namin et al. (2010) presented a sensitivity reduction due to the irregular shapes of the nodules candidates. Our work uses extraction methods fully based on texture to overcome this situation. Lee et al. (2010b) and Pu et al. (2008a) their approaches utilize techniques that require parameterization, empirically obtained. The original AC method was adapted in this work to not need any parameterization, as well as RD method. Finally, Tartar et al. (2013) and Zhang et al. (2013) methods are adapted to extract features from two-dimensional slices from the CT exams. This may reduce the natural characteristics of the lung nodules, if the chosen slices do not com-

prise all information necessary to recognize the object. This work uses all the three-dimensional properties of the CT scans in both proposed techniques.

Table 1 summarizes the approaches.

These studies contributed to the construction of the proposed methodology. A number of techniques are explored and the testing methods we use are identical to those of Carvalho Filho et al. (2014). These methods were chosen because they are rigorous and yielded reliable results.

3. Background

This section describes the techniques used in the feature extraction step of both the artificial crawlers and the rose diagram.

3.1. Artificial crawlers

Artificial Life are models that imitate life in nature Langton (1989). The individuals that compose these models attempt to behave in similar ways to real organisms, and their behaviours, actions and lifespan are abstracted in a logical form. Based on the Artificial Life concept, evolutionary algorithms and ACs were originally designed as models that use two-dimensional images as the environment. The individual, called the artificial crawler (AC), lives in the pixels of grey-level images (Zhang & Chen, 2004).

Throughout evolution, some actions (movement, improvement, competition, stability, survival and death) are observed in the AC individuals. Spatially close individuals are labelled as participants of the identical colony. The colonies, as well as the numbers of participants, change as evolution occurs.

Each AC has certain characteristics: t is the time or iterator of the ACs algorithm, $e(t)$ is the energy of the individual in time t ; $\delta(t)$ is the colony the individual belongs to in time t ; and $\beta(t)$ is the location - the pixel - the AC is set in time t .

The objective of each AC is survival. They move to the nearest pixel with the highest intensity (local maximum) as they absorb energy from the environment to achieve this goal. The higher the intensity of the pixel, the greater the amount of energy absorbed.

There are seven rules that the ACs must follow. The rules are applied to every time t until the end of the cycle for every AC i .

1. $\forall_i, e_i(0) = e^0$. The initial value of every AC is e^0 .
2. \forall_i, t , if $e_i(t) < e_{min}$, then AC $_i$ dies. At each cycle, if the energy of any AC is less than a minimum value e_{min} , it dies. The AC is removed from the image.
3. $\forall_i: e_i(t) \geq e_{min}, \beta_i(t+1) = f(\beta_i(t))$. At each time t , the ACs will move according to three possible conditions, presented in Eq. (1).

$$f(\beta) = \begin{cases} \beta & \text{if (a)} \\ \beta_{max} & \text{if (b)} \\ \beta' & \text{if (c)} \end{cases} \quad (1)$$

- (a) If the maximum pixel intensity of the 8-pixel neighbourhood of $\beta_i(t)$ of the AC i is lower than the pixel intensity of $\beta_i(t)$, the AC does not move.
- (b) If the maximum pixel intensity of the 8-pixel neighbourhood of $\beta_i(t)$ of the AC i is higher than the pixel intensity of $\beta_i(t)$ and it is unique - i.e., there is only one pixel in the neighbourhood with higher intensity -, the AC i moves to that location.
- (c) If the maximum pixel intensity of the 8-pixel neighbourhood of $\beta_i(t)$ of the AC i is higher than the pixel intensity of $\beta_i(t)$ and it is not unique - i.e., there is more than one pixel in the neighbourhood with higher intensity -, the AC i moves to a pixel that has already been occupied by another AC.
4. $\forall_i: e_i(t) \geq e_{min}, e_i(t+1) = e_i(t) - e_{unit}$. When an AC i moves, it will lose one unit of energy defined by e_{min} .

Table 1

Summary of the related works.

Works	Objective	Database	Cases	Key differences
Mousa and Khan (2002)	Nodule classification	Proprietary	50	Circular features and mean variance to classify lung nodules and non-nodules using SVM
Jing et al. (2010)	Nodule classification	LIDC	50	Lung nodule classification using SVM and shape, geometry, grey level and texture features
Lee et al. (2010b)	Nodule classification	LIDC	32	Classification aided by clustering using random forest
Namin et al. (2010)	Nodule segmentation and classification	LIDC	63	FKNN classification using texture and geometry features
Tartar et al. (2013)	Nodule classification	Proprietary	63	Lung nodules classification using PCA and mrMR features
Zhang et al. (2013)	Nodule classification	ELCAP	50	Lung nodule classification using superpixel labelling and context curve
Choi and Choi (2014)	Nodule classification	LIDC	84	Nodule and non-nodule classification using angular histograms of the normal surface
Carvalho Filho et al. (2014)	Nodule classification	LIDC-IRDI	833	Nodule Classification and non-nodule structures using taxonomic diversity indices
Pu et al. (2008a)	Lung segmentation	Proprietary	20	Adaptive borders algorithms to delimit lung
Shen et al. (2015)	Nodule segmentation	LIDC	233	Bidirectional chain-code and SVM for classification
Tan et al. (2011)	Nodule segmentation	LIDC	235	Invariant features extraction and SVM and Artificial Neural Network for classification
Pu et al. (2008b)	Nodule detection	Proprietary	52	Progressive clustering and Shape similarity analysis for lung nodule detection

5. $\forall_{i,j}: \beta_i(t+1) = \delta_j(t+1), \quad \sigma_{\text{sigmax}_{e_i,e}} = \max\{e_i, e_j\}$ and $\text{sigmax}\{e_i, e_j\}$ dies. When ACs i and j occupy the same location at the same time, the one with the higher energy survives and the other one dies.
6. $\forall_i, e_i(t+1) = e_i(t) + \lambda I(x, y)$. The AC i absorbs λ percent of the pixel intensity it lies in at time t , if it moved. This amount is converted into energy.
7. $\forall_{i,j}: e_i(t) \geq e_{\max}, e_i(t+1) = e_{\max}$. The maximum energy the ACs can retain is defined as e_{\max} .

The rules are applied to all ACs at all times. We can observe the state of emergency and balance of the population. With evolution, no individual will die or regroup in a colony (emergency) at a certain time. Certain ACs will move to areas of high intensity to absorb energy, losing a unit simultaneously (balance). The result is a state of balance.

By recording the steps taken in the evolution of the ACs, we can later extract useful information. The numbers of iterations and colonies in the ACs algorithm reach equilibrium.

These properties characterize the texture features of the environment through a series of evolution curves. The curves are: agent evolution, inhabitant settlement, colony formation and scale distribution (Zhang et al., 2013).

The agent evolution curve represents the number of ACs per epoch. The number of ACs drops at each iteration because of the depletion of energy, and depends on the placement of valleys and peaks of intensity in the environment.

The second curve, inhabitant settlement, is similar to the agent evolution curve. It represents the number of extant ACs per iteration.

The curve of colony formation depicts the number of colonies formed by elements within a radius of distance. The increment of the radius size results in fewer colonies with more individuals per colony.

The scale distribution depicts the statistical representation of the colonies at different scales. Different textures lead to varied scale distributions. This curve is an important technique for discerning texture.

3.2. Rose diagram

The RD is a circular histogram with bars that represent sectors (Mardia & Jupp, 2009). The RD resembles a flower, and the sectors resemble petals. The number of sectors in the RD is determined by the user and the group angle intervals.

The RD extracts information based on angular features of objects. Many directional measurements, such as local gradients, can be used. The Sobel filter is commonly used to calculate the gradients (Lee, Seo, Ryoo, & Yoon, 2010a).

Some features may be extracted using the Sobel gradients (Mardia, 1972). One of these features is the mean direction of the gradient angles (X), which can be calculated using Eq. (2).

$$X = \arctan\left(\frac{\sum_{i=1}^n \sin\theta_i}{\sum_{i=1}^n \cos\theta_i}\right) + \lambda \quad (2)$$

in which θ is the representative angle of the sector,¹ n is the total number of angles and λ is represented by Eq. (3).

$$\lambda = \begin{cases} \pi & \text{if } \sum_{i=1}^n \sin\theta_i > 0 \\ 0 & \text{if } \sum_{i=1}^n \cos\theta_i > 0 \text{ and } \sum_{i=1}^n \sin\theta_i > 0 \end{cases} \quad (3)$$

The complement of the mean direction, which is called the circular variance (S_0), is computed using Eq. (4).

$$S_0 = 1 - X \quad (4)$$

The circular standard variation (s_0) is calculated by Eq. (5):

$$s_0 = \sqrt{-2 * \log_e(1 - S_0)} \quad (5)$$

in which S_0 is the circular variance.

The mean resulting length (F) is computed using Eq. (6):

$$F = \sqrt{\frac{\sum_{i=1}^n \sin^2\theta_i + \sum_{i=1}^n \cos^2\theta_i}{n}} \quad (6)$$

in which n is the total number of angles.

The kurtosis (K) is computed by Eq. (7):

$$K = \frac{1}{n} * \sum_{i=1}^n \cos(2 * (\theta_i - X)) \quad (7)$$

in which X is the mean direction of the gradient angles and n is the total number of angles.

The skewness (A) can be extracted from the RD and is computed using Eq. (8).

$$A = \frac{1}{n} * \sum_{i=1}^n \sin(2 * (\theta_i - X)) \quad (8)$$

¹ Each rose diagram sector groups a range of angles. The larger the range, the larger the size of each section, and the total number of sections is decreased.

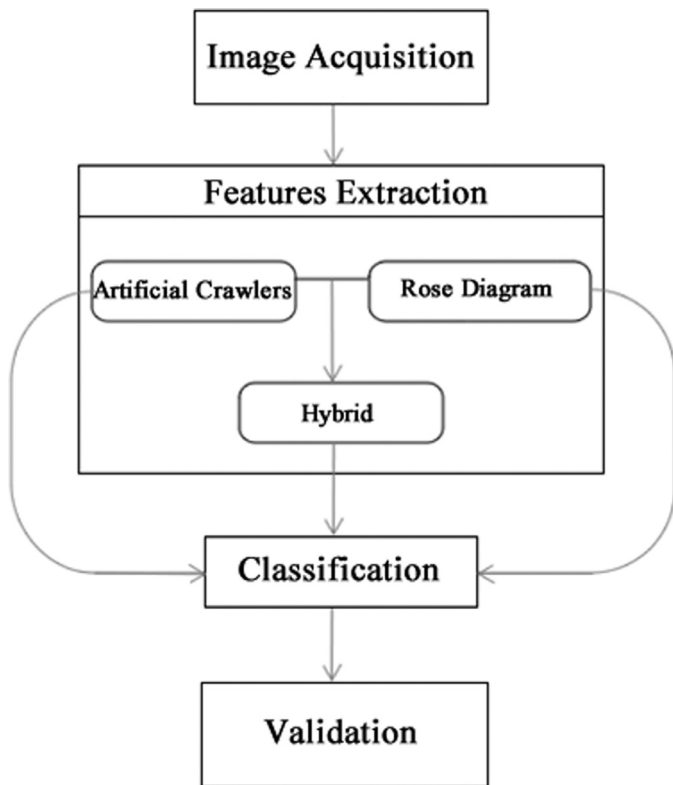


Fig. 1. Proposed methodology flowchart.

in which X is the mean direction of the gradient angles and n is the total number of angles.

4. Proposed methodology

In this section, we describe the proposed methodology. The main goal of this work is to construct a consistent methodology for the classification of candidates into nodules or non-nodules. There are four phases: image acquisition, feature extraction, classification and validation. In the first phase, nodule candidates are extracted from LIDC-IRDI and the non-nodule candidates are extracted from [Carvalho Filho et al. \(2014\)](#). The second phase is feature extraction, which is performed using three methods: artificial crawlers, the rose diagram and a hybrid model from both models. The third phase is the classification using the SVM classifier. The last phase is the validation of the classification, intended to measure the efficacy of the methodology. [Fig. 1](#) illustrates the proposed methodology.

4.1. Image acquisition

The lung images used in this work are available in the public database LIDC-IRDI ([Armato et al., 2011](#)). Some principles, such as identity protection, patient consent and lung cancer diagnosis, are defined by the publisher.

The database contains 1012 CT scan series and 290 radiography scan series. Each image contains information regarding the nodules, such as location, size, quantity and malignancy. The scans are diagnosed by four specialists and all diagnostics are indicated in the auxiliary files. Some scans contain conflicting information between the image header and the auxiliary files. Therefore, only 833 CT scans were chosen to be part of the candidate database. The images are divided into two databases: nodule and non-nodule candidates.

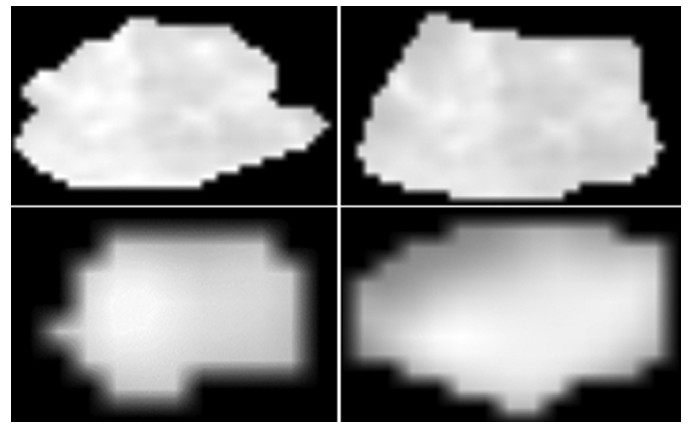


Fig. 2. 2D view of nodule candidates.



Fig. 3. 2D view of non-nodule candidates.

A CT scan, provided by the LIDC-IRDI database, contains lung nodules markers from the four specialists. The scan may have one or more nodules marked individually. Sometimes the specialists disagree on the diagnosis. These markers contain the coordinates and the slices of the 3D CT exam in which the nodules are found. The size of the nodule varies from 3 mm to 30 mm. The database does contain nodule markers that are less than 3 mm, but they are represented by a single coordinate, not by a bounding box. These smaller nodules are not included in this work. [Fig. 2](#) depicts some examples of nodule candidates used in the proposed methodology.

We utilize all regions marked by specialists as nodules. For example, if the image has four markings, we use them in the training and test as four distinct nodules. The idea is to use all the specialist expertise.

The non-nodule candidates were extracted from the methodology described by [Carvalho Filho et al. \(2014\)](#). Their work uses the quality threshold algorithm followed by a growing region to segment lung nodule and non-nodule candidates from the LIDC-IRDI database.

We used non-nodule candidates from [Carvalho Filho et al. \(2014\)](#) because their work is consistent. Non-nodules are excluded from the detection step and can be reused in the proposed methodology. Their nodule candidate database is not employed in the present work because the database is small (and may thus result in misclassification) compared to the specialist markers from LIDC-IRDI. [Fig. 3](#) demonstrates examples of non-nodule candidates.

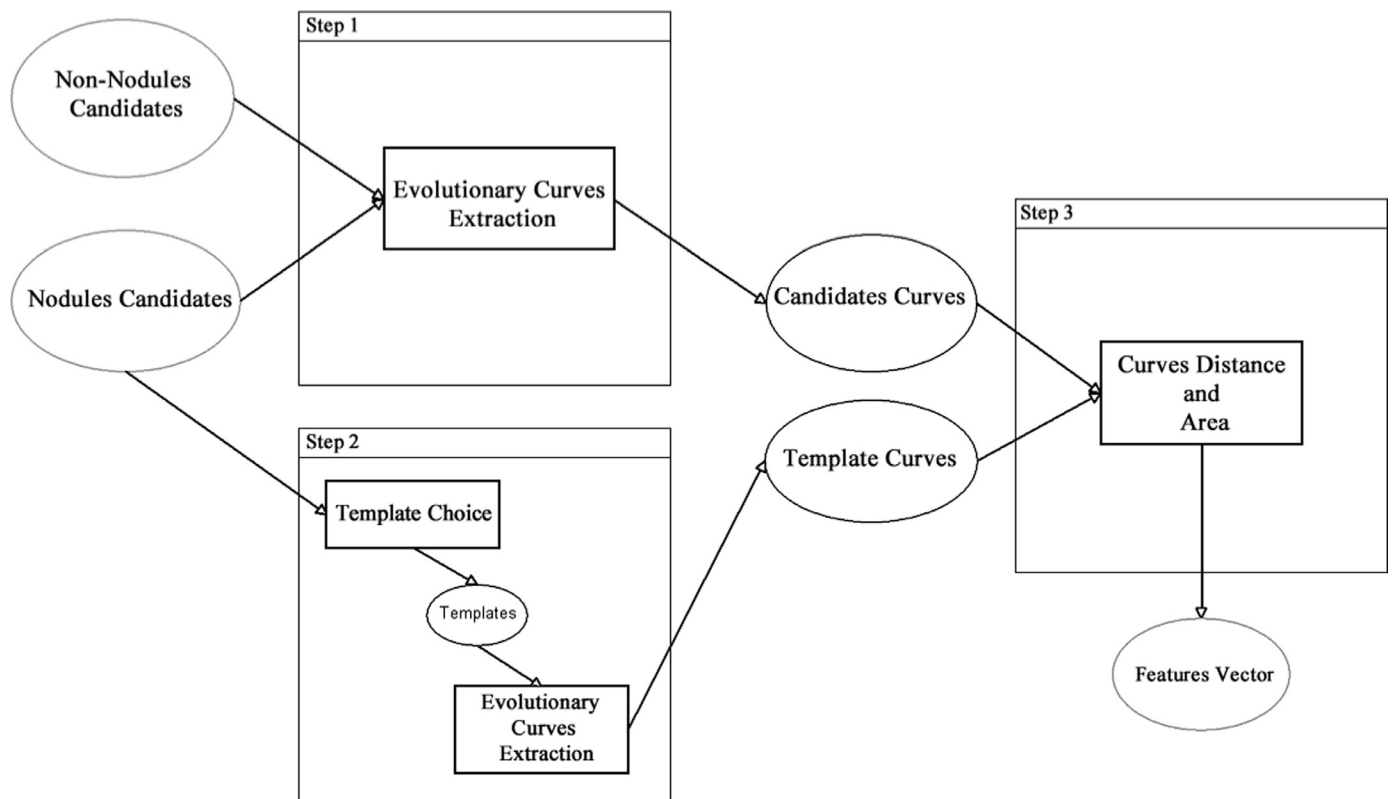


Fig. 4. ACs model diagram.

4.2. Feature extraction

The extraction of features occurs in this phase. This is a necessary step for posterior classification. The proposed methodology uses three techniques for feature extraction. The first one is the ACs model, the second is the RD model, and the third combines features from the ACs and the RD.

Directional texture (the rose diagram model) exploits the advantages of both the directional and the intensity information in the images (Rivera, Castillo, & Chae, 2015). Analyzing that information in a local neighbourhood diminishes the influence of noise in the feature extraction because the gradient works with the edges, regardless of single pixel intensity variation.

The use of artificial life concept in image classification resulted in promising results by Zhang and Chen (2004). The ACs method provided high level statistical features that represent consistencies in the textures of the images.

Both techniques utilize different types of analyses inside the texture. If used together, they may reduce weaknesses and the classifier has the advantages of both methods.

4.2.1. Artificial crawlers

The ACs model and the steps necessary to extract the feature vectors are presented in Fig. 4.

Step 1 uses both nodule and non-nodule candidates and retrieves the corresponding evolutionary curves from the ACs algorithm for all images. In step 2, the templates are chosen to aid in posterior feature extraction, which is based on the differences between the curves. The templates of nodule candidates are the CT scans that best represent the entire database. The more similar the curves of the nodule templates are to the candidate analyzed, the higher the probability of the candidate belonging to the nodule class. Step 3 uses the differences between the candidates and

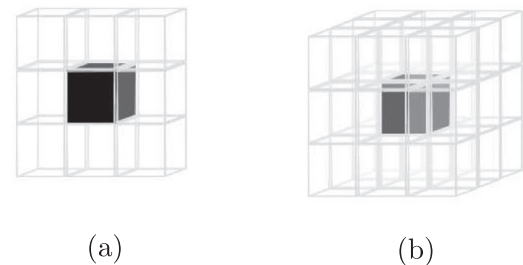


Fig. 5. An AC's perceived neighbourhood. The black cube is the AC's voxel. The transparent cubes are the neighbourhood voxels. (a) 2D 8-pixel neighbourhood from Zhang and Chen (2004). (b) The proposed 3D 26-pixel neighbourhood.

the template curves and the area under the candidate curves as features.

Step 1: The extraction of the evolutionary curves (the ACs algorithm is applied to every candidate from the two separate databases) generates four evolutionary curves.

We added some modifications to the ACs method developed by Zhang and Chen (2004). We adapted the ACs to CT scans because they are 3-dimensional images. ACs individuals may move in the x, y and z axes. The ACs perceive the 26-neighbourhood space instead of the previous 8-neighbourhood plane. Fig. 5 demonstrates the AC adaptation for this work.

The new ACs assign the number of initial ACs using the voxels not in the background as the location for the new-born individuals. The old ACs approach assigns random positions to individuals. This approach is not effective for lung nodule and non-nodule candidate images because elements born in background voxels will not interact with texture variations.

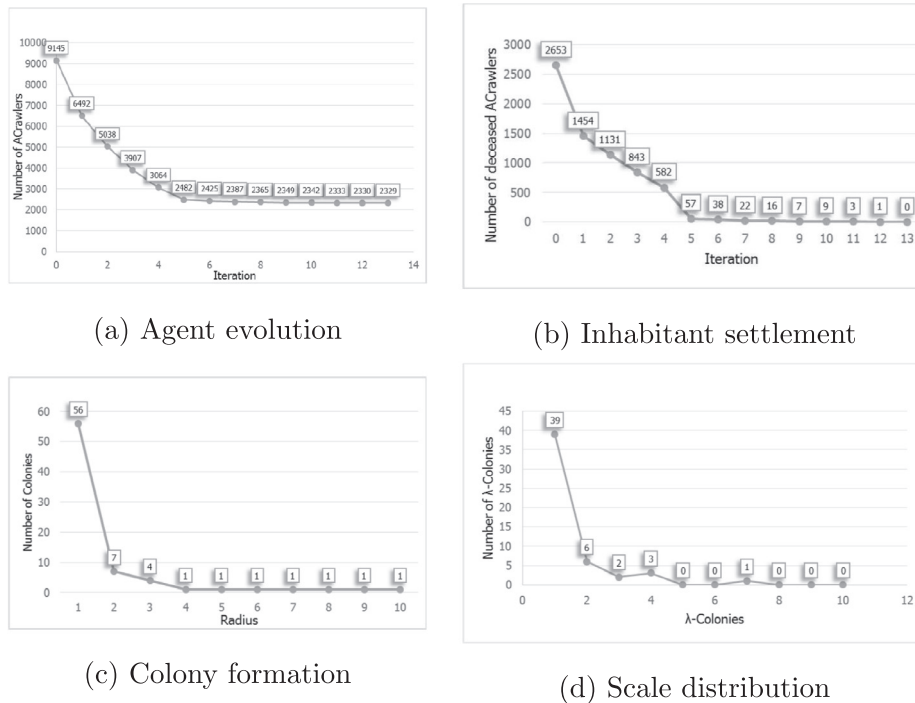


Fig. 6. Curves extracted from one nodule candidate.

Individuals will be isolated and the curve extraction will contain irrelevant information regarding the environment. We can observe the behaviour of each candidate curve at the end of the extraction. Fig. 6 depicts the curves extracted from one nodule candidate.

Step 2: In this step, one or more nodule candidates are chosen as templates. Templates represent the entire nodule candidate database for further comparison between the templates and other candidates.

We choose templates only from the nodule candidates because of the inferior number of candidates in this database. It is easier to find elements that best represent the entire set using a smaller base.

The nodule candidate database contains extra information regarding the size of the nodules, which is used to facilitate the choice of the template. We chose three templates (3 mm to 10 mm, 11 mm to 20 mm and 21 mm to 30 mm) because of the variation in the size of the nodules. After choosing the template, the evolutionary curves are extracted.

Step 3: The final step creates the feature vector. With the evolutionary curves of the candidates and the templates obtained, the features are calculated in two ways: curve differences and area under the curve.

The curve distance will calculate the similarity between all candidates and all templates individually. The sizes of the images in the database are different because of the diversity of the candidates in the samples. This diversity may lead the ACs algorithm to create a variable number of initial ACs. The distance between curves may vary too much, even when curves acquired from similar textures have identical behaviour. The candidate curves are normalized between 0 and 1 in the y-axis (maintaining the shape of the curves) to reduce this inconvenience.

The distances used are the Euclidean (Danielsson, 1980), Jaccard (Shameem & Ferdous, 2009), simple matching (Paragios, Rousson, & Ramesh, 2002), Chebyshev

(Klove, Lin, Tsai, & Tzeng, 2010) and Manhattan (Chang, Desoky, Ouyang, & Rouchka, 2009). For each template, the five distances are calculated and put into the features vector. The area under the evolutionary curve of each candidate is included in the feature vector.

4.2.2. Features from rose diagram model

The RD model is presented in Fig. 7. The model contains the method diagram and the steps necessary to extract the feature vector.

In step 1, the Sobel filter is applied to every candidate, generating the derivatives G_x and G_y , a pair (G_x, G_y) for each candidate image. These pairs of derivatives are used in step 2 for the computation of the gradients. The Sobel gradients are analyzed through the RD perspective in order to extract the features and to create the feature vector.

Step 1: The Sobel filter is applied to the candidates. One pair of derivatives (G_x, G_y) is obtained for each candidate (Fig. 8). The model contains the method diagram and the steps necessary to extract the feature vector.

This operation is executed in every slice of a candidate because it is a two-dimensional method. The gradients are grouped together into a single RD per candidate.

Step 2: The Sobel gradients are computed for every pair of derivatives (G_x, G_y) in all non-background image pixels. There is no texture variation in background pixels, so it is irrelevant to use the gradient information at these locations.

Step 3: The last step uses all gradients and creates one RD perspective per candidate. Fig. 9 depicts one RD from the gradients computed from the identical nodule candidate to Fig. 8a. The following features are calculated from the populated RD: mean direction, circular variance, circular standard variation, mean resulting length, kurtosis and skewness. These calculations are entered into the feature vector.

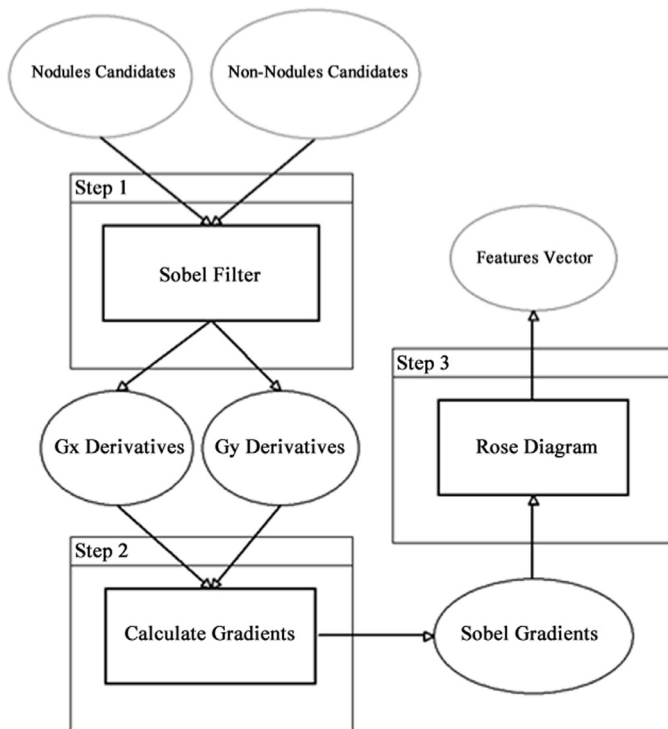


Fig. 7. RD model diagram.

Table 2
Features list.

	Artificial crawlers	Rose diagram
Features	agent evolution inhabitant settlement colony formation scale distribution curves areas	mean direction circular variance circular standard variation mean resulting length kurtosis skewness

4.2.3. Features from hybrid model

The hybrid model is composed of the ACs and RD models. It results in one feature vector which is a combination of the feature vectors produced by the other two models. Fig. 10 depicts the hybrid model scheme.

Table 2 summarizes all features present in the ACs and RD models. These features will be combined into a single feature vector in the hybrid model.

4.3. Classification

The nodule and non-nodule candidates are treated as two separate classes. We used the support vector machine (SVM) classifier with the features extracted in Section 4.2.

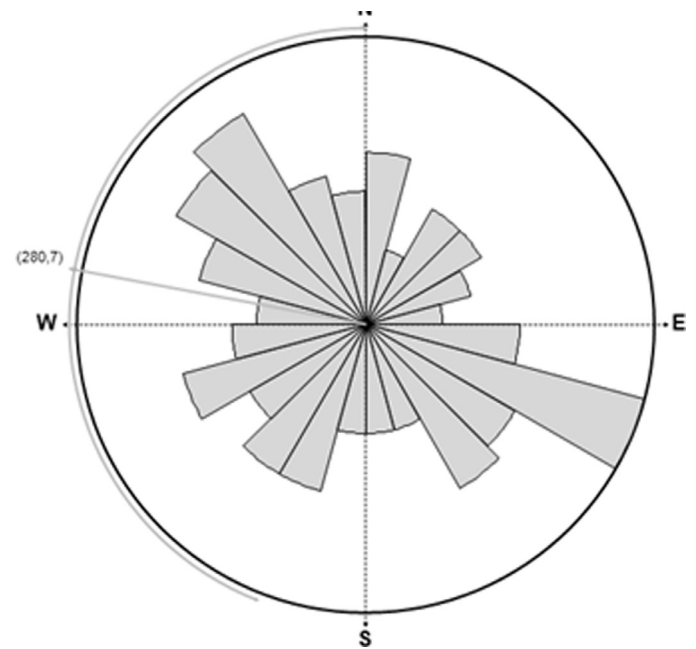


Fig. 9. RD of a nodule candidate (Fig. 8a).

SVM is a group of supervised learning algorithms used for classification or regression Joachims (1998). It is straightforward to change kernels with SVM. This property allows the description of complex data relationships.

An implemented version of SVM, the open source library LIB-SVM (Chang & Lin, 2011), is used in this work through the tool Weka 3.6. The chosen kernel is the RBF, and the parameters c and γ are estimated using the grid search algorithm (Bergstra & Bengio, 2012).

4.4. Validation

The last phase demonstrates the consistency of the methodology, measuring its predictive value. It computes the accuracy, specificity, sensitivity (Fielding & Bell, 1997), variation coefficient of accuracy (Soliman, Abd Ellah, Abou-Elheggag, & Abd-Elmougod, 2011) and receiver operating characteristic (ROC) curve (Bradley, 1997).

5. Results and discussion

The entire database of candidate images is partitioned into two groups: training and test. The features are extracted using the three proposed methods. The results are evaluated and compared to other related works.

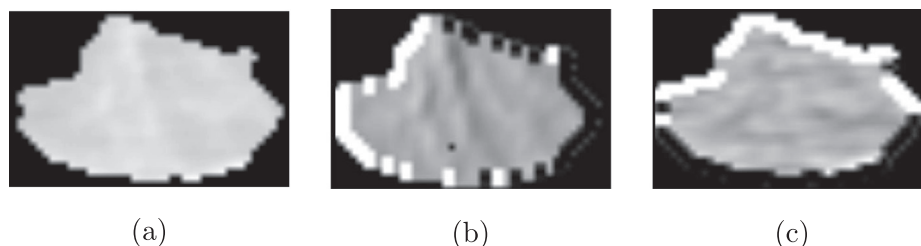


Fig. 8. Sobel filter applied to one nodule candidate slice. (a) The original image (b) Gx derivative (c) Gy derivative.

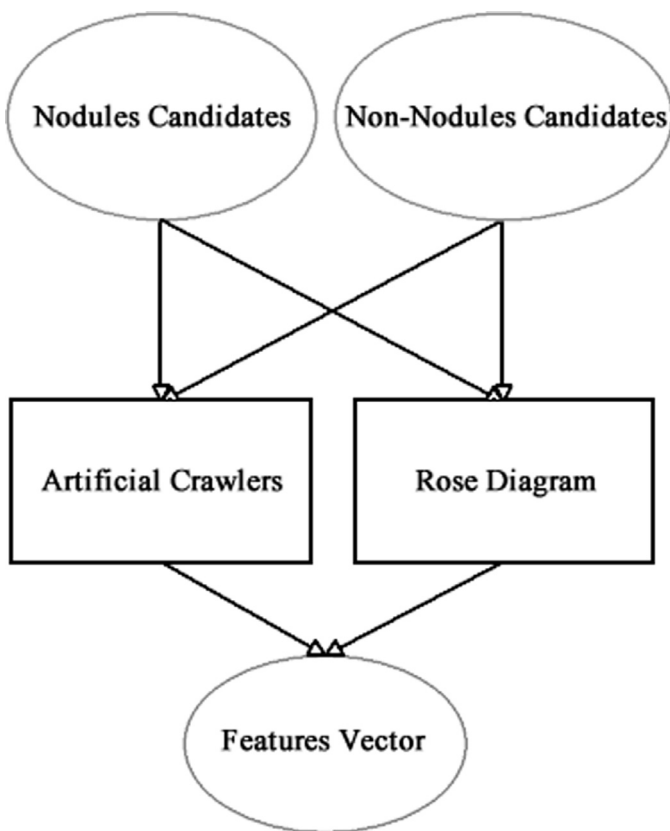


Fig. 10. Hybrid model diagram.

5.1. Database separation

We used the LIDC-IDRI (Armato et al., 2011) image database which is available on the Internet as a result of the association between the Lung Image Database Consortium and the Image Database Resource Initiative. All images are in DICOM format and have up to 16 bits per voxel. The computed tomographies (CT) were acquired using various protocols, which complicates the detection of lung nodules.

In this database, each exam is assigned to a file containing information regarding the contour and features of each nodule, supplied by four specialist physicians. The contour indicates the points (x,y,z) of the nodule. The entire region inside the contour is taken into consideration during the evaluation by our methodology. The features indicate the properties of the nodule, including: degree of difficulty at detection, calcification level, sphericity level, degree of spiculation and degree of malignancy. However, this information is only valid for nodules with diameters larger than 3 mm and smaller than 30 mm. For nodules smaller than 3 mm, the specialists indicate only the centre of the mass. Regions larger than 30 mm are considered to be masses.

The LIDC-IDRI contains 1012 CT scans. Some (179) scans contain two factors that make our methodology impossible. The first factor is related to scans that have no nodules larger than or equal to 3 mm. These files do not contain the properties cited above. The second factor is related to divergences in the marking file, when it does not match the information contained in the DICOM header. Therefore, we only applied the methodology to the remaining 833 scans.

We used the 833 scans as training and testing samples for classification. We extracted 6415 nodule candidates from these scans based on the specialists' markings. From Carvalho Filho et al. (2014), we retrieved 17,742 non-nodule candidates. Even with



Fig. 11. Sample slices.(a) Nodule and (b) Non-nodule example.



Fig. 12. Comparison of non-nodule and nodule agent evolution curves.

nearly three times as many non-nodules as nodules, the SVM classifier is able to optimize the separation of these two datasets due to the distinct characteristics vectors between the two samples.

The candidate database was proportionally divided into training and testing subgroups. Four types of proportions were used: 20% for training and 80% for testing, 40% for training and 60% for testing, 60% for training and 40% for testing and 80% for training and 20% for testing. Each proportion is permuted five times. There are five distinct distributions for the candidate training/testing proportions. The goal of this approach is to demonstrate that the results are consistent.

All training/testing groups provide the mACC, the mSEN, the mSPC, the CAV and the mROC of the five distributions evaluated for each proportion.

5.2. Curves analysis

The slope of the evolution curves extracted from the ACs method can be interpreted in several ways. Fig. 11 depicts two samples of slices, a nodule and a non-nodule. We compare the curves extracted from these two sample slices.

The agent evolution curve illustrates the number of living ACs per iteration. Curves with a steep initial downward trend represent the fast ascending of ACs to high intensity locations. The non-nodules do not have a specific behaviour because there is no texture pattern related to high and low intensity. However, the nodules normally display high densities in a few locations because of calcification. Fig. 12 shows a visual representation of this curve with the two sample slices from Fig. 11.

The inhabitant settlement curve has a smooth behaviour, similar to the nodule agent evolution curve. Images with locations of diffuse high intensity may have more deceased ACs for the initial cycles. However, the number decreases in subsequent iterations. This phenomenon is common for non-nodules. Fig. 13 demonstrates the behaviour of the curves in both cases.

We can interpret the colonies with smaller radius as diffuse peaks with high intensities isolated and spatially distant through the texture of the image. Curves with a high number of colonies

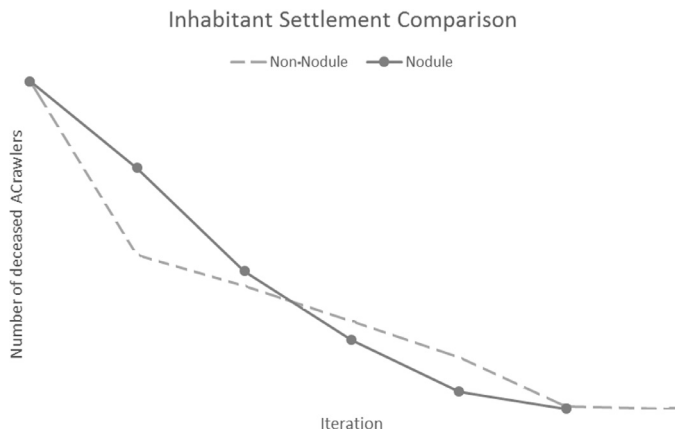


Fig. 13. Comparison of non-nodule and nodule inhabitant settlement curves.

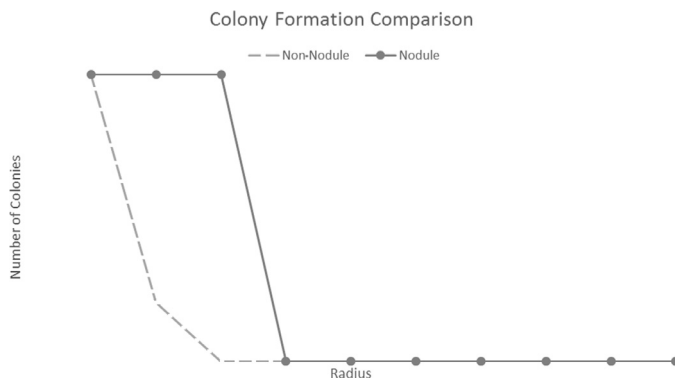


Fig. 14. Comparison of non-nodule and nodule colony formation curves.

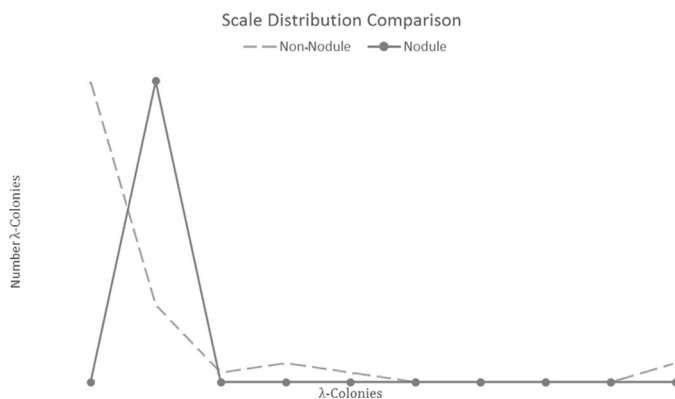


Fig. 15. Comparison of non-nodule and nodule scale distribution curves.

with a small radius are found at the final process of ACs in non-nodule samples. Fig. 14 demonstrates this behaviour in comparison to the slope of the colony formation curve for nodules.

The scale distribution of non-nodules is irregular. There are large groups of ACs at the end of the ACs process. This behaviour is uncommon in the scale distribution of nodules because few ACs remain close to one another. This phenomenon is because of the low number of peaks with high intensity. Fig. 15 depicts the slope of both cases of scale distribution.

5.3. Tests

The features are extracted using the three previously described methods. The first extraction is provided by the ACs method. The

Table 3

Results obtained by the ACs method.

Training/testing %	mACC %	mSEN %	mSPC %	CAv %	mROC
20/80	93.28	91.48	93.82	1.39	0.901
40/60	93.4	91.82	93.9	1.67	0.909
60/40	92.46	91.74	93.62	2.18	0.905
80/20	93.27	91.34	93.38	2.5	0.908

Table 4

Results obtained through the RD method.

Training/testing %	mACC %	mSEN %	mSPC %	CAv %	mROC
20/80	64.18	65.24	65.04	7.71	0.651
40/60	65.82	56.38	71.02	5.31	0.637
60/40	66.9	61.82	70.46	4.33	0.661
80/20	66.72	68.26	66.96	2.14	0.676

Table 5

Results obtained through the hybrid method.

Training/testing %	mACC %	mSEN %	mSPC %	CAv %	mROC
20/80	93.54	90.58	94.46	1.5	0.908
40/60	94.2	91.7	94.84	1.14	0.918
60/40	94.3	91.86	94.78	1.61	0.922
80/20	94.22	92	94.52	1.94	0.921

second extraction uses the RD method and the third extraction is performed using the hybrid model.

The classification was performed using the SVM classifier. Each method was tested separately and the best results are compared to the literature and discussed.

The parameters estimated through the grid search technique to the RBF kernel of SVM were c equal to 512 and γ equal to 2.

The results of the ACs method for feature extraction are presented in Table 3. The best result achieved by this method was with the proportion of 40% for training/60% for testing, with mACC of 93.4%, mSEN of 91.82%, mSPC of 93.9%, CAv of 1.67% and mROC of 0.909.

The results achieved with RD are reported in Table 4. The best results were observed with the proportion of 60% for training/40% for testing. The values obtained are lower than the ACs method, with mACC of 66.9%, mSEN of 61.82%, mSPC of 70.46%, CAv of 4.33% and mROC of 0.661.

The results achieved with the hybrid method are listed in Table 5. The best results were achieved in this approach with a proportion of 60% for training/40% for testing. We report mACC of 94.3%, mSEN of 91.86%, mSPC of 94.78%, CAv of 1.61% and mROC of 0.922.

The methodology proposed in this work demonstrated consistent results for the nodule classification problem. There was a notable discrepancy between the results obtained in tests between the AC and RD models. However, the results were improved by the combination of these models.

The performances of the methods imply some very useful information. The test cases separated by different proportions showed very similar results in each model. That means the randomly chosen partitions are well balanced even with more non-nodules candidates than nodules candidates.

The RD method performs better on true negative rate (mSPC). Its power to correctly designate true negatives candidates is clearly observed on Hybrid method, because in comparison to AC model, the mSPC was improved. This is the main contribution of the RD features on Hybrid method.

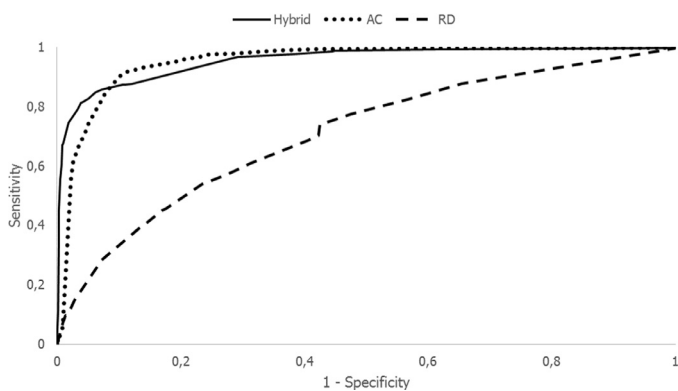


Fig. 16. mROC curve visualization from 60%/40% training/testing results.

Table 6
Computational costs for training step.

Method	Mean elapsed time (seconds)		
	Feature extraction	Model evaluation	Total
ACs	13,764	47.64	13811.64
RD	358.46	129.14	487.6
Hybrid	14122.5	215.98	14338.48

Table 7
Computational costs for testing stage.

Method	Mean elapsed time (seconds)		
	Feature extraction	Model testing	Total
ACs	8.2	0.5	8.7
RD	0.3	0.25	0.55
Hybrid	8.5	0.7	9.2

The best results were achieved with the hybrid model with a training/testing proportion of 60%/40%. The worst results were from the RD model tests. The RD model approach is not able to classify lung nodules consistently. However, the combination of the ACs and RD models is more reliable than the two models alone.

The best result for the mROC curve can be observed in Fig. 16. The mROC curve demonstrates the good performance of the hybrid method in comparison with the ACs and RD approaches. Any increase in sensitivity will be accompanied by a decrease in specificity. We observe that the optimal true positive rate is achieved near the lowest value of the false positive rate.

As can be observed from the hybrid mROC curve, the hybrid method performance benefits mainly from the ACs method. However, the RD method also benefits the hybrid performance, al-

though with less influence. The reason for this is the adaptation of the SVM classifier with a radial basis in scenarios with characteristics of undetermined relevance. The most beneficial features are also the most relevant ones. So the classifier is responsible for using the relevant features of both methods to optimize the hybrid performance.

Tests of the methodology were implemented on an Intel i5-3230M CPU @2.60 GHz on-board personal computer. The total average execution times of the extraction of features, training and testing steps were recorded and are presented in Tables 6 and 7.

Table 6 reports the mean average time consumed during the training stage for each method. The feature extraction is the most time-consuming stage. The hybrid method is the most time-consuming because it combines the ACs and RD methods.

This stage uses the entire database to retrieve the necessary information for feature extraction and the classification for model evaluation. This evaluation is performed using the SVM classifier and the extracted features.

The training stage can be performed only one time and is recorded in a file for later testing.

Table 7 reports the mean elapsed time, in seconds, for the testing stage. This is the least time-consuming stage because it is performed on only one sample. With the trained database, the feature extraction and the model testing (classification) are fast enough to fit in a CAD system with acceptable performance.

5.4. Comparison with related works

Table 8 lists the results achieved in the literature. These results can be compared to the results obtained using the proposed methodology. The comparison is approximate because some works used different databases and test scenarios.

Our results are obtained from the hybrid model in the proportion of 60% for training and 40% for testing. The only accurate comparison is with Carvalho Filho et al. (2014). Their methodology was evaluated with the identical conditions.

Compared to the work by Carvalho Filho et al. (2014), our methodology achieved a smaller specificity. It is less capable of correctly classifying the non-nodule cases. Sensitivity measures the capability of correct classification in nodule cases (patients with the illness) and our methodology is more efficient. Achieving a good balance between the sensitivity, specificity and accuracy is important. This balance was not found in Carvalho Filho et al. (2014). It is preferable for the methodology to have the smallest number of errors in the cases with cancer.

Other works use smaller or proprietary databases, which reduces the reliability of the results in comparison to the results in the present study. Inasmuch some works use different study cases, we implemented three classical techniques for texture analysis us-

Table 8
Comparison between our results and results of related works.

Works	Database	Cases	Sensitivity %	Specificity %	Accuracy %	ROC
Mousa and Khan (2002)	Proprietary	50	87.5	–	–	–
Jing et al. (2010)	LIDC	50	–	–	85.44	–
Lee et al. (2010b)	LIDC	32	98.33	97.11	–	0.9786
Namin et al. (2010)	LIDC	63	–	–	88	–
Tartar et al. (2013)	Proprietary	63	89.6	87.5	90.7	–
Zhang et al. (2013)	ELCAP	50	–	–	82.5	–
Choi and Choi (2014)	LIDC	84	97.2	97.7	97.4	–
Carvalho Filho et al. (2014)	LIDC-IRDI	833	85.91	97.7	97.55	0.8062
Pu et al. (2008a)	Proprietary	20	–	–	90	–
Shen et al. (2015)	LIDC	233	–	–	92.6	–
Tan et al. (2011)	LIDC	235	87.5	–	–	–
Pu et al. (2008b)	Proprietary	52	95.1	–	–	–
proposed work	LIDC-IRDI	833	91.86	94.78	94.30	0.922

Table 9

Comparison between classical techniques and proposed method.

Method	Accuracy %	Sensitivity %	Specificity %
Histogram	81.91	72.4	84.44
Gray-Level Co-Occurrence Matrix	84.26	75.54	86.18
Gray Level Run Lengths	85.34	81.55	86.56
Proposed method (Hybrid)	94.30	91.86	94.78

ing our tests scenarios for a righteous comparison of the proposed method. The first one is the Histogram, the second one is the Gray-Level Co-Occurrence Matrix and the last one is the Gray Level Run Lengths. The results are presented on Table 9.

Through Table 9, it can be observed that the proposed methodology is superior to the classical methods, with the same database, classification method and evaluation conditions. The proposed work demonstrates great potential with this comparison.

6. Conclusions

This work presented a method that uses ACs, RD and SVM for classifying lung nodules and non-nodules.

The tests performed with the hybrid model yielded 94.3% mean accuracy, 91.86% mean sensitivity, 94.78% mean specificity, 1.61% coefficient of variation of the accuracy and 0.922 mean area under the ROC curve. These results were obtained using the proportion of 60% training and 40% testing.

Some adaptations were made to the ACs model. The first adaptation enables the use of CT scans as the environment. The original ACs methodology used only 2D images. The other change is the elimination of the random factor in image location for newborn ACs. In the RD model, the use of a Sobel gradient as directional data extracted from the texture of the candidates was designed. The directional statistical measurements were extracted using this method.

The changes to the ACs model are the main contribution of this work. These changes adapted the model for the classification of medical images.

Although the good results presented in this article, we believe this methodology can be improved. For this purpose, we propose the following as future work:

1. The methodology uses the original intensity values of the voxels. However, related works use some techniques to accentuate the images qualities. That means the classification results should improve if any pre-processing method fits well in the methodology.
2. The Rose Diagram method use fixed size for the sectors to extract the gradient features. An adaptive way to choose the size of the sector may optimize the method result.
3. The templates used by AC method used the nodule size for defining the templates. Several characteristics can be used to determining templates, i.e. the nodules malignancy features. This can embrace new opportunities for others classification goals, like the distinctions between benign and malign nodules.
4. There can be used related methods in the artificial life context. Other evolutionary algorithms and genetic algorithms may be used to confront or support the AC method for better performance.
5. This approach may also be adapted for temporal series images of Cancer CT exams for recognize nodules and its behavior over time, a four-dimension analysis.

Both the ACs and RD methods were not widely used in the nodule classification problem. The evaluation of both techniques, and the creation of a hybrid method represent an important contribution to the scientific community.

Acknowledgements

The authors acknowledge CAPES, CNPq and FAPEMA for their financial support. The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health for their critical role in the creation of the free, publicly available LIDC-IDRI database used in this study.

References

- Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, 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. *Medical Physics*, 38(2), 915–31. Accessed 08.11.15.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305. Accessed 08.11.15.
- Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. doi:10.1016/S0031-3203(96)00142-2.
- Carvalho Filho, A. O., de Sampaio, W. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., & Gattass, M. (2014). Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial Intelligence in Medicine*, 60(3), 165–177. doi:10.1016/j.artmed.2013.11.002.
- Chang, C.-C., & Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 27:1–27:27. doi:10.1145/1961189.1961199.
- Chang, D.-J., Desoky, A. H., Ouyang, M., & Rouchka, E. C. (2009). Compute pairwise manhattan distance and pearson correlation coefficient of data points with gpu. In H.-K. Kim, & R. Lee (Eds.), *Software engineering, artificial intelligences, networking and parallel/distributed computing*, 2009. snpd'09. 10th acis international conference on (pp. 501–506). Daegu, Korea. IEEE: IEEE Computer Society. doi:10.1109/SNPD.2009.34.
- Choi, W.-J., & Choi, T.-S. (2014). Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Computer Methods and Programs in Biomedicine*, 113(1), 37–54. doi:10.1016/j.cmpb.2013.08.015.
- Danielsson, P.-E. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing*, 14(3), 227–248. doi:10.1016/0146-664X(80)90054-4.
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(01), 38–49. Accessed 08.11.15.
- Jing, Z., Bin, L., & Lianfang, T. (2010). Lung nodule classification combining rule-based and svm. In K. Li, Z. Tang, R. Li, A. K. Nagar, & R. Thamburaj (Eds.), *Bio-inspired computing: Theories and applications (bic-ta)*, 2010 IEEE fifth international conference on (pp. 1033–1036). Tasmania, Australia: IEEE. doi:10.1109/BICTA.2010.5645114.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In C. Nédellec, & C. Rouveirol (Eds.), *Machine learning: Ecml-98. In Lecture Notes in Computer Science: 1398* (pp. 137–142). Springer Berlin Heidelberg. doi:10.1007/BFb0026683.
- Klove, T., Lin, T.-T., Tsai, S.-C., & Tzeng, W.-G. (2010). Permutation arrays under the chebyshev distance. *Information Theory, IEEE Transactions on*, 56(6), 2611–2617. doi:10.1109/TIT.2010.2046212.
- Langton, C. G. (1989). *Artificial life: The proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems: 6*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.. Accessed 08.11.15.
- Lee, H., Seo, S., Ryoo, S., & Yoon, K. (2010). Directional texture transfer. In M. McGuire, & J. Collomosse (Eds.), *Proceedings of the 8th international symposium on non-photorealistic animation and rendering* (pp. 43–48). New York, NY, USA. ACM: ACM. doi:10.1145/1809939.1809945.
- Lee, S. L. A., Kouzani, A. Z., & Hu, E. J. (2010). Random forest based lung nodule classification aided by clustering. *Computerized Medical Imaging and Graphics*, 34(7), 535–542. doi:10.1016/j.compmedimag.2010.03.006.
- Mardia, K. V. (1972). *Statistics of directional data* (pp. 18–38). Academic Press. doi:10.1016/B978-0-12-471150-1.50008-X.
- Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics*: 494. John Wiley & Sons. Accessed 08.11.15.
- Mousa, W. A., & Khan, M. A. (2002). Lung nodule classification utilizing support vector machines. In B. Mercer (Ed.), *Image processing. 2002. proceedings. 2002 international conference on*: 3 (pp. 153–156). New York, USA: IEEE. doi:10.1109/ICIP.2002.1038927.
- Namin, S. T., Moghaddam, H. A., Jafari, R., Esmaeil-Zadeh, M., & Gity, M. (2010). Automated detection and classification of pulmonary nodules in 3d thoracic CT images. In *Proceedings of the IEEE international conference on systems, man and cybernetics, Istanbul, Turkey, 10–13 October 2010* (pp. 3774–3779). Istanbul, Turkey. doi: 10.1109/ICSMC.2010.5641820.
- Paragios, N., Rousson, M., & Ramesh, V. (2002). Matching distance functions: A shape-to-area variational approach for global-to-local registration. In A. Heyden, G. Sparr, M. Nielsen, & P. Johansen (Eds.), *Computer vision—eccv 2002. In Lecture Notes in Computer Science: 2351* (pp. 775–789). Springer Berlin Heidelberg. doi:10.1007/3-540-47967-8_52.

- Polakowski, W. E., Cournoyer, D. A., Rogers, S. K., DeSimio, M. P., Ruck, D. W., Hoffmeister, J. W., et al. (1997). Computer-aided breast cancer detection and diagnosis of masses using difference of gaussians and derivative-based feature saliency. *Medical Imaging, IEEE Transactions on*, 16(6), 811–819. doi:[10.1109/42.650877](https://doi.org/10.1109/42.650877).
- Pu, J., Roos, J., Yi, C. A., Napel, S., Rubin, G. D., & Paik, D. S. (2008). Adaptive border marching algorithm: Automatic lung segmentation on chest ct images. *Computerized Medical Imaging and Graphics*, 32(6), 452–462. doi:[10.1016/j.compmedimag.2008.04.005](https://doi.org/10.1016/j.compmedimag.2008.04.005).
- Pu, J., Zheng, B., Leader, J. K., Wang, X.-H., & Gur, D. (2008). An automated ct based lung nodule detection scheme using geometric analysis of signed distance field. *Medical Physics*, 35(8), 3453–3461. doi:[10.1118/1.2948349](https://doi.org/10.1118/1.2948349).
- Rivera, A. R., Castillo, J. R., & Chae, O. (2015). Local directional texture pattern image descriptor. *Pattern Recognition Letters*, 51, 94–100. doi:[10.1016/j.patrec.2014.08.012](https://doi.org/10.1016/j.patrec.2014.08.012).
- Shameem, M.-U.-S., & Ferdous, R. (2009). An efficient k-means algorithm integrated with jaccard distance measure for document clustering. In G. Omidyar (Ed.), *Internet, 2009. ah-ici 2009. first asian himalayas international conference on* (pp. 1–6). Kathmandu, Nepal: IEEE. doi:[10.1109/AHICI.2009.5340335](https://doi.org/10.1109/AHICI.2009.5340335).
- Shen, S., Bui, A. A., Cong, J., & Hsu, W. (2015). An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Computers in Biology and Medicine*, 57, 139–149. doi:[10.1016/j.combiomed.2014.12.008](https://doi.org/10.1016/j.combiomed.2014.12.008).
- Soliman, A., Abd Ellah, A., Abou-Elheggag, N., & Abd-Elmougod, G. (2011). A simulation-based approach to the study of coefficient of variation of gompertz distribution under progressive first-failure censoring. *Indian Journal of Pure and Applied Mathematics*, 42(5), 335–356. doi:[10.1007/s13226-011-0022-8](https://doi.org/10.1007/s13226-011-0022-8).
- Tan, M., Deklerck, R., Jansen, B., Bister, M., & Cornelis, J. (2011). A novel computer-aided lung nodule detection system for ct images. *Medical Physics*, 38(10), 5630–564. doi: [10.1118/1.3633941](https://doi.org/10.1118/1.3633941).
- Tartar, A., Kilic, N., & Akan, A. (2013). Classification of pulmonary nodules by using hybrid features. (pp. 1–11). (vol. 2013). doi:[10.1155/2013/148363](https://doi.org/10.1155/2013/148363).
- Wood, D. E., Eapen, G. A., Ettinger, D. S., Hou, L., Jackman, D., Kazerooni, E., et al. (2012). Lung cancer screening. *Journal of the National Comprehensive Cancer Network*, 10(2), 240–265. Accessed 08.11.15.
- Zhang, D., & Chen, Y. Q. (2004). Classifying image texture with artificial crawlers. In N. Zhong, J. Bradshaw, S. K. Pal, D. Talia, J. Liu, & N. Cercone (Eds.), *Intelligent agent technology, 2004.(iat 2004). proceedings. ieee/wic/acm international conference on* (pp. 446–449). Beijing, China: IEEE. doi:[10.1109/IAT.2004.1342992](https://doi.org/10.1109/IAT.2004.1342992).
- Zhang, F., Song, Y., Cai, W., Zhou, Y., Shan, S., & Feng, D. (2013). Context curves for classification of lung nodule images. In P. de Souza, U. Engelke, & A. Rahman (Eds.), *Digital image computing: Techniques and applications (dicta), 2013 international conference on* (pp. 1–7). Hobart, Tasmania: IEEE. doi:[10.1109/DICTA.2013.6691494](https://doi.org/10.1109/DICTA.2013.6691494).