Journal of Advances in Computer Research
Quarterly ISSN: 2008-6148
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 5, No. 1, February 2014), Pages: 19-28
www.jacr.iausari.ac.ir

Detection of Pulmonary Nodules in CT Images Using Template Matching and Neural Classifier

Hosien Hasanabadi[1 *], Mohsen Zabihi[2], Qazaleh Mirsharif[3]

1) Department of Computer Engineering, Quchan Branch, Islamic Azad University, Quchan, Iran
2) Department of Computer Engineering, Ferdowsi University, Mashhad, Iran
3) Department of Computer Engineering, Shiraz University, Shiraz, Iran

hosien.hasanabadi@yahoo.com; zabihi@um.ac.ir; mirsharif@shiraz.ac.ir

Received: 2013/09/02; Accepted: 2013/10/28

Abstract

Computer aided pulmonary nodule detection has been among major research topics lately to help the early treatment of lung cancer which is the most lethal kind of cancer worldwide. Some evidence suggests that periodic screening tests with the CT of patients will help in reducing the mortality rate caused by the lung cancer. A complete and accurate computer aided diagnosis (CAD) system for detection of nodules in lung CT images consists of three main steps: extraction of lung parenchyma, candidate nodule detection and false positive reduction. While precise segmentation of lung region speed upthe detection process of pulmonary nodules by limiting the search area, in candidate nodule detection step we attempt to include all nodule like structures. However, the main problem in the current CAD systems for nodule detection is the high false positive rate which is mostly associated to misrecognition of juxta-vascular nodules from blood vessels. In this paper we propose an automated method which has all of the three above mentioned steps. Our method attempts to find initial nodules by thresholding and template matching. To separate false positives from nodules, we use feature extraction and neural classifier. The proposed method has been evaluated against several images in LIDC database and the results demonstrate improvements in comparison with the previous methods.

Keywords: *False positive reduction; neural classifier; pulmonary nodule detection; template matching*

1. Introduction

Lung cancer has been reported as the leading cause of cancer related death around the world. Lung cancer often does not develop symptoms until it is in its advanced stage.Therefore, early detection and diagnosis of lung nodules has became a focus of research to raise the survival chance of patients suffering from lung cancer.

Computed tomography (CT) of the patients (particularly thin section scans) has proved as the best imaging modality for characterization of the small pulmonary nodules. However, analysis of the large number of scans makes it a cumbersome task for radiologists to manually assess the tumor biology. As a results several CAD systems have been developed lately aimed at providing radiologists with a second opinion to estimate the size and location of pulmonary nodules.

An accurate CAD system for pulmonary nodule detection consists of three main steps: lung region extraction, candidate nodule detection and false positive reduction. Precise extraction of lung region and elimination of noise and unrelated structures will speed up the detection of candidate nodules by limiting the area where nodules are to be sought and next calculations are applied to. For identification of initial nodules several methods such as multi-thresholding, Gaussian filters, and match filters have been proposed in the literature. Most of the recently presented CAD systems for pulmonary nodule detection have high sensitivity rate. The main problem in the current systems is that the number of their false positives is still high. Since the nodules are frequently attached to blood vessels or pleura, many non-nodule anatomical structures are recognized as nodules. To reduce the false positive rate, a third step is usually added which attempts to classify candidate nodules to nodules and false positives.

In this paper we have proposed a CAD system for pulmonary nodule detection which consists of the three mentioned steps. In our method, we use thresholding and template matching for initial nodule detection. Also we employ a neural classifier for false positive reduction.

The rest of the paper is organized as follows: Section II reviews the methods proposed for pulmonary nodule detection in CT images. In section III LIDC dataset which has been used in this paper is introduced. In section IV the proposed method is explained in details. In section V the results are discussed and finally section VI concludes the paper.

2. Related Work

Although CAD has been applied to various image modalities, CT has provided better results due to its sensitivity. Particularly in thin-section CT scans of patients, small nodules are more likely to be detected, therefore we only review methods that have been applied to thin-section CT images.

In 2007 Qiang Li [1] reported the performance of CAD systems for lung cancer detection in thin-section CT images. In a thin section CT scans, the section thickness is small typically between 1 to 2.5 mm. A CAD scheme can generally detect and characterize small and possibly curable cancers more reliably in thin section CT than in thick-section CT, because the partial-volume effect is much lower in thin section CT than in thick-section CT, and also because three-dimensional (3D) image processing and analysis techniques become applicable in thin-section CT.

Qiang Li [1] described a general CAD system by two major phases including suspicions nodule identification and false positive reduction. Some methods such as Li et al;[2] however involve a nodule enhancement processing which helps in detection of low contrast nodules and those connected to blood vessels or airway walls. They showed superior results comparing to methods that do not include an enhancement on nodules [3,4]. Bae et al;. [5] Developed a method for detection of nodules in three categories: isolated, jaxtapleural and juxtavascular nodules. They employed spherical shape morphological filters to enhance juxtavascular nodules.

Many methods used single or multi-thresholding for initial identification of nodules directly from the images without applying nodule enhancement [6,7,8].Ge and Gerun et al. [4,9] employed weighted k-means clustering segmentation for initial nodule identification including a feature extraction and classification process.

For false positive reduction step, candidate nodules must be classified into nodules and non nodules. Many investigators [10 ,11 ,12] have used a rule-based classifier to distinguish nodules from false positives. Because the rule-based classifier generally has a clear semantic meaning, it can be readily comprehended or interpreted by human beings. However, rules were generally determined manually and empirically in existing CAD schemes, which leads to tediousness, long design time, and an overtraining effect. Also neural networks, linear discriminant analysis and genetic algorithms have been used for differentiation of nodules from false positives [11,13 ,14].

Also in a recent work Gin Mo Goo [15] studied status of research on available CAD for pulmonary nodule detection. A wide range of sensitivities from 55 to 95% with false positive rates of 0.55 to 8.3 per examination were reported in this study. However, it is difficult to compare the results of CAD's performance from the various studiesas most of the performances were evaluated with different data sets, and particularly a different size database, the different nature and characteristics of the nodules and various evaluation methods, as well as the differences in the employed CAD algorithm.

Although studies have shown that CAD increased the observer's performance in detection of pulmonary nodules, giving the moderate sensitivities, it still acts as a second reader.

In this paper we proposed a method which attempts to improve the sensitivity while decreasing the number of false positives. We evaluated our method on public dataset of LIDC which is introduced in the next section.

3. DataBase

The LIDC/IDRI Database [16] contains 1018 cases, each of which includes images from a clinical thoracic CT scan and an associated XML file that records the results of a two-phase image annotation process performed by four experienced thoracic radiologists. In the initial blinded-read phase, each radiologist independently reviewed each CT scan and marked
lesions belonging to one of three categories ("nodules >=3 mm," "nodule<3 mm," and "non-nodule>=3 mm""). In the subsequent unblinded-read phase, each radiologist independently reviewed their own marks along with the anonymized marks of the three other radiologists to render a final opinion. The goal of this process was to identify as completely as possible all lung nodules in each CT scan without requiring forced consensus.

The Database contains 7371 lesions marked "nodule" by at least one radiologist. 2669 of these lesions were marked "nodule>=3 mm" by at least one radiologist, of which 928 (34.7%) received such marks from all four radiologists. These 2669 lesions include nodule outlines and subjective nodule characteristic ratings.

4. Method

In this section the proposed method for pulmonary nodule identification is explained in details. The presented method consists of three main steps. First, the lung region where nodules are located must be extracted. This task is performed using thresholding, edge detectors and morphological filters. In the second step, we attempt to find candidate nodules in two phases using thresholding in consecutive slices and template matching method. The combined results from this step are further analyzed in the final

step for elimination of false positives. Ultimately, features are extracted from initial nodules and candidate nodules are classified into nodules and false positives using a neural classifier.

A. Extraction of pulmonary parenchyma

The extraction methodology of the pulmonary parenchyma can be regulated by removing the background as an initial step (i.e., the pixels with the same grey level as the lungs but located outside the chest) from the image to avoid confusion.

This is intended, due to the high similarity between the grey levels of the lungs and the image background, which cannot be simply applied by using a Thresholding technique. Instead an ad-hoc operator is used which, starting from the four corners of the image, moves along the four directions identifying as background pixels, those pixels whose grey level is within a pre-fixed range. Figure 1 shows an original image from CT before removing the back ground.

When a pixel value is found outside the range (out pixel), then it is analysed for few more pixels in the scan direction. If any of the pixels have values within the range then both these pixels and the out pixel are marked as background pixels and the scan goes on to be continued. If not, the scan is interrupted along the direction under examination and the successive row or the particular column is analyzed in a similar way. These images produced by the operator are converted to binary images by means of some specific technique like Thresholding technique that uses either a static or a dynamic threshold depending on the lung zone and the slice that belongs to the zone in the lungs.

More specifically, after discarding the first few initial and last slices of a CT scan that do not contain lung images the remaining slices are classified into three groups corresponding to the upper, middle and lower parts of the lung volume. In the first and last groups the lung image represents a smaller percentage of the slice than in the second group. Therefore a different thresholding technique is applied depending on the group the given slice belongs to. By using a dynamic threshold determined through an iterative procedure for the slices of the middle part of the lung, a threshold is determined empirically for all other slices. The thresholding operators produce each slice which is transformed into a binary image where the foreground, i.e., the object we are interested in, consists of the lungs and all the rest is background. See figure 1.

Afterwards, the morphological opening and closing operators are applied so as to improve the image and border definition to enhance the separation between distinct regions and to fill the gaps in the borders.

Then we detect the image borders through a tracking algorithm that uses the Sobel operator and reduce the border

size to one pixel using thinning algorithm. In this algorithm, a

pixel is considered a border pixel if at least one of its neighbors is white. The algorithm starts building all the border pixels chains and then eliminates the excess pixels producing a one-pixel contour image.

B. Candidate nodule detection

Lung nodule detection is a very difficult step in a CAD system development. Actually, in CT lung images, nodules are frequently attached to blood vessels or to the pleura; further their grey tone is so similar to vessels sections that traditional intensity-based methods are inappropriate. Instead, an effective nodule detection algorithm must

take both the greylevel and the object shape into account. As a result, a method that attempts to detect nodules on a slice by slice basis may not produce good results.

Radiologists indeed trace suspicious structures in a sequence of slices to see if they are nodule or part of ananatomical structure which belong to lung. If a detected structure with higher gray value than background is vessel it will appear in consecutive slices while a nodule is of a spherical limited shape and it is not of fixed gray value in different slices.

We use a two phase process for identification of candidate nodules. We first apply a threshold to lung regions to eliminate black pixels. Also morphological filters are employed to remove the imaging noise. The threshold is also applied to the ten previous and next slices of the current image. Number of slices for which a suspicious structure passes the threshold value is more for a vessel than for a nodule. Moreover noise will disappear very fast by thresholding. This is indeed a fuzzy co-occurrence of gray values for pixels of a nodule or vessel.

However we should note that when a nodule is attached to vessels, it might be considered as a part of vessel and will be removed. To overcome this problem, nodules must be separated from vessels. Therefore each image is divided to small windows. As a result, juxtavascular nodules will not be eliminated completely. Unfortunately, dividing the image into windows and applying the threshold to each window will also lead to elimination of small isolated nodules.

We perform the nodule candidate detection process in two parallel stages and combine their results. First, the initial threshold is applied to the lung region to separate the parts of lung with higher gray values including nodules, noise and anatomical structures as explained above. In the first stage we follow the pixels with higher values than background in previous and next slices to separate the nodules from noise and vessels. In the second stage, we employ a template matching algorithm which is applied to each image after initial thresholdingin which regions with higher gray value has been detected. By combining the results from these two separate phases, we attempt to retrieve suspicious nodules.

Nodules are of different sizes, shapes and gray values. In LIDC lung images nodules are marked and separated from non nodules. Different nodule patterns have been collected from LIDC database for detection of nodules in CT images. See figure 3. These 60 nodule patterns have been averaged to help the extension of the method to other datasets of lung images. Using similarity measurement, similarity between each nodule pattern and detected regions is measured. Afterwards, those pixels for which their similarity measurement for at least one pattern passes the threshold are marked.

Figure 1. Lung parenchyma extraction

To measure the similarity we use:

$$sim(k) = \frac{\sum_{i,j} p(i,j) - temp_k(i,j)}{\sqrt{\sum_{i,j} p^2(i,j) - temp^2(i,j)}} \quad (1)$$

Where P is a window and tempk is the kth pattern.

The results combined from these two phases are the candidate nodules which include nodules and false positives. See figure 2. There are still some non-nodule structures that have been mistakenly identified as nodules. To minimize the false positives, we add another step to the proposed method. In this final step we attempt to minimize the number of false positives using feature extraction from pixels of candidate nodules. These two last steps are applied lung region identified by a segmentation algorithm which defines the internal region [17].

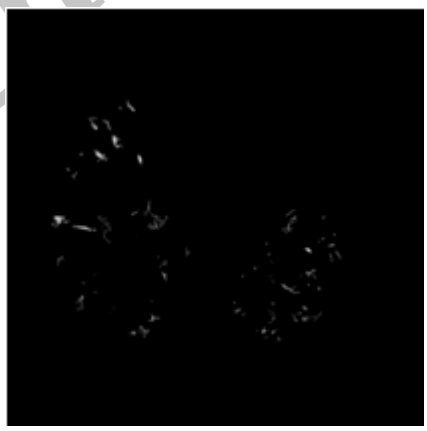

Figure 2. candidate nodules

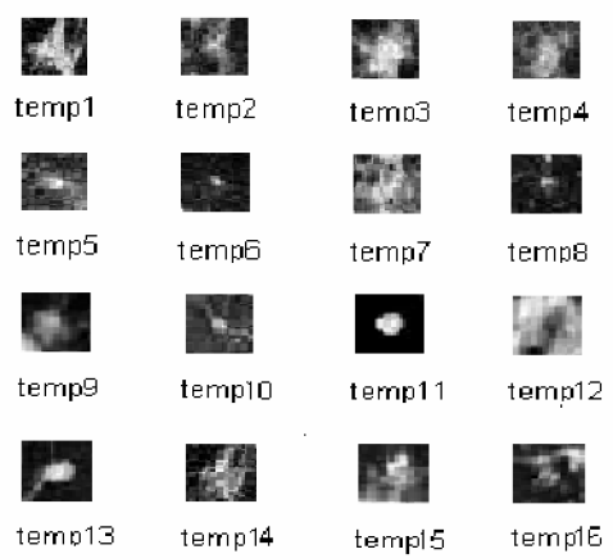

Figure 3. some of the nodule patterns extracted from LIDC

C. False positive reduction

Available cad systems for lung nodule detection mostly have high false positive rate. For false positive reduction features are usually extracted from candidate nodules and they are classified to nodules and false positives.
Most false positive findings are crossings between blood vessels. To reduce the amount of FP per scan, voxel-based neural approach (VBNA) is developed [13].

Each voxel of a region of interest (ROI) is characterized by a feature vector constituted by the grey level intensity values of its 3D neighbors (Figure 4) and the eigen values of the gradient and the Hessian matrices [18]. A feed-forward neural network is trained and tested by assigning each voxel either to the nodule or normal tissue target class. Finally a candidate nodule is characterized as "CAD nodule", in this case if the number of pixels within its ROI is labeled as "nodule" by the neural classifier, it exceeds some qualified threshold.

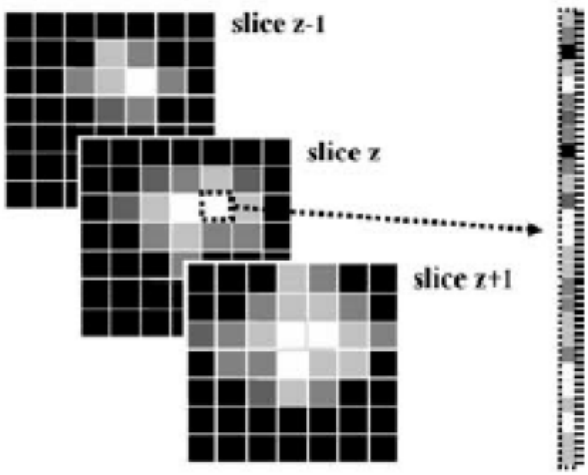

Figure 4. Voxel based neural approaches for reduction of false positive

5. Results and discussion

The proposed CAD has been applied to CT images of 7 patients from LIDC database which includes 1100 slices. All the images have been preprocessed and lung region is extracted for each of them. The CT for each patient contains different number of nodules. The results of the presented CAD system are demonstrated in terms of sensitivity and false positive per scan. Since the candidate nodule detection step includes all nodules that have been marked by radiologists, the sensitivity of the method is 100%. In the final step false positive rate is reduced. The final results for each patient are seen in Table I. See figure 5.For each patient there are several types of nodules. The proposed method detects nodules attached to vessels or bronchial walls as well as solid nodules. Yet, most false positives detected by the method are associated to attached nodules.

Table II shows the results obtained by other methods. We should note that the presented method in the literature has been evaluated against different datasets of lung images with different number and types of nodules. Precise comparison of methods is not possible unless the methods are tested on the same set of data. Also it has been shown that we can increase sensitivity of methods in the cost of increasing the false positive numbers. However we attempt to increase the sensitivity while keeping the false positive rate of the method low. The approach employed for candidate method is fast and easy to implement due to using various nodule patterns.

Yet we could improve the method by evaluating against different set of data particularly with a focus on detection of nodules attached to anatomical structures in lung which cause the major part of false positive rate.

Table 1. Results of the proposed classification method

Image number	Number of slices	Numder of slices with nodules	Number of nodule types	Sensitivity	False positive
13614193285030025	269	5	4	100%	1
13614193285030026	268	39	23	100%	2
13614193285030027	136	17	24	100%	3
13614193285030030	104	23	10	100%	2
13614193285030031	109	13	23	100%	3
13614193285030032	106	18	70	100%	4
13614193285030033	116	11	13	100%	3

Table 2. Results obtained by other methods

method	DataBASE	sensitivity	false positive rate
BAE [5,19]	20 PATIENTS WITH 164 SOLID NODULES	91.2%	6.9 FPS/SCAN
BROWN[7]	29 PATIENTS 77 NODULES	70%	15 FPS/SCAN
PAIK[10]	8 PATIENTS	80%	1.3 FPS/SCAN

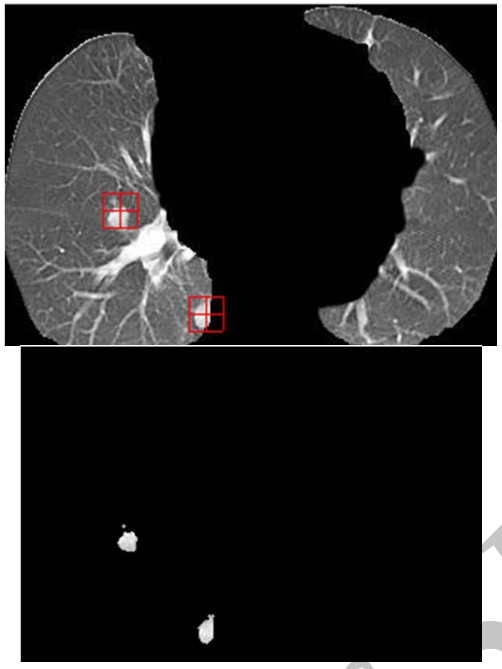

Figure 5. Results of the proposed CAD on one example image

6. Conclusion

In this paper a CAD system for lung nodule detection has been proposed. The proposed method consists of three main steps including lung parenchyma extraction,candidate nodule detection and false positive reduction.For candidate nodule identification we use a two phase method. First a threshold is applied to all the images to separate regions with higher gray values than background.These regions are then followed in consecutive slices to detect nodules from other structures in lung. Also a template matching approach is employed to find the similarity between the detected regions and 60 nodule patterns collected from LIDC dataset. Combined results from these two phases constitute candidate nodules. Finally in the third step features are extracted from candidate nodules and they are classified into nodules and false positives by a neural classifier.The proposed CAD for pulmonary nodule detection has been evaluated against several images in LIDC database and the results show progress comparing to previous methods.

For the future we suggest to evaluate the method on different set of images with a concentration on identification of nodules attached to vessels or walls of lung which cause the major part of false positives. Currently CAD is employed as a second reader to assist radiologists for lung cancer detection. An enhancement on the available methods may be obtained by investigation of juxtavascular nodules in different datasets of thin-section CT scans of patients.

Acknowledgments. The authors would like to thank officials in Islamic Azad University, Quchan Branch for their financial support of the scientific project.

7. References

[1] Li Q, "Recent Progress in Computer-Aided Diagnosis of Lung Noduleson Thin-Section CT,"*Comput Med Imaging Graph*. 2007 ; 31(4-5): 248–257.

[2] Li Q, Sone S, Doi K. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. Med Phys 2003;30:2040–2051. [PubMed: 12945970]

[3] Lee Y, Hara T, Fujita H, Itoh S, Ishigaki T. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. IEEE Trans Med Imag 2001;20:595–604.

[4] Gurcan MN, Sahiner B, Petrick N, Chan HP, Kazerooni EA, Cascade PN, Hadjiiski L. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Med Phys 2002;29:2552–2558. [PubMed: 12462722]

[5] Bae KT, Kim JS, Na YH, Kim KG, Kim JH. Pulmonary nodules: Automated detection on CT images with morphologic matching algorithm – Preliminary results. Radiology 2005;236:286–294. [PubMed: 15955862]

[6] Boroczky L, Zhao L, Lee KP. Feature subset selection for improving the performance of false positive reduction in lung nodule CAD. IEEE Trans Info Tech in Biomedicine 10:504–511.

[7] Brown MS, Goldin JG, Suh RD, McNitt-Gray MF, Sayre JW, Aberle DR. Lung micronodules: automated method for detection at thin-section CT – initial experience. Radiology 2003;226:256–262. [PubMed: 12511699]

[8] Zhao BS, Gamsu G, Ginsburg MS, Jiang L, Schwartz LH. Automatic detection of small lung noduleson CT utilizing a local density maximum algorithm. J Applied Clin Med Phys 2003;4:248–260.[PubMed: 12841796]

[9] Ge ZY, Sahiner B, Chan H-P, Hadjiiski LM, Cascade PN, Bogot N, Kazerooni EA, Wei J, Zhou C. Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field methodand 3D ellipsoid fitting. Med Phys 2005;32:2443–2454. [PubMed: 16193773]

[10] Paik DS, Beaulieu CF, Rubin GD, Acar B, Jeffrey RB, Yee J, Dey J, Napel S. Surface normal overlap:A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. IEEE Trans Med Imag 2004;23:661–675.

[11] Lawler LP, Wood SA, Pannu HS, Fishman EK. Computer-assisted detection of pulmonary nodules: preliminary observations using a prototype system with multidetector-row CT data sets. J Digital Imaging 2003;16:251–261.

[12] Marten K, Engelke C, Seyfarth T, Grillhosl A, Obenauer S. Computer-aided detection of pulmonary nodules: Influence of nodule characteristics on detection performance. Clin Radiology 2005;60:196–206.

[13]Gomathi M, Thangaraj T. Lung Nodule Detection using a Neural Classifier. IACSIT International Journal of Engineering and Technology, Vol.2, No.3, June 2010

[14] Ye X, Lin X, Dehmeshki J. Shape Based Computer-Aided Detection of LungNodules in Thoracic CT Images. IEEE Trans Biomed Eng. 2009 Jul;56(7):1810-20.

[15] Jin Moo Goo, A Computer-Aided Diagnosis for Evaluating LungNodules on Chest CT: the Current Status and Perspective. Korean J Radiol 2011;12(2):145-155

[16]G.Armato S, et al. The Lung Image Database Consortium „LIDC… and Image Database Resource Initiative (IDRI): A Completed Reference Databaseof Lung Nodules on CT Scans. Med. Phys. 38(2)February 2011

[17] J. Heuberger, A. Geissbühler, and H. Müller, "Lung CT segmentationfor image retrieval using the Insight Toolkit (ITK)," Medical Imagingand Telemedicine (MIT), August 2005.

[18] I. Gori, M. Mattiuzzi, "A method for coding pixels or voxels of a digitalor digitalized image: Pixel Matrix Theory (PMT)," European PatentApplication 05425316.6.

[19] Kim JS, Kim JH, Cho GS, Bae KT. Automated detection of pulmonary nodules on CT images: Effect of section thickness and reconstruction interval – Initial results. Radiology 2005;236:295–299.