Editorial Manager(tm) for International Journal of Computer Assisted Radiology and Surgery

Manuscript Draft

Manuscript Number: JCARS-D-10-00163R2

Title: Combination of Computer-Aided Detection algorithms for automatic lung nodule identification

Article Type: Original Article

Keywords: Computer Aided Detection; Computed Tomography; Lung Cancer; Medical Image Analysis; Pattern Recognition

Corresponding Author: Niccolò Camarlinghi

Corresponding Author's Institution: Dipartimento di Fisica dell'Università di Pisa and Istituto Nazionale di Fisica Nucleare Sezione di Pisa, Pisa, Italy

First Author: Niccolò Camarlinghi

Order of Authors: Niccolò Camarlinghi;Ilaria Gori;Alessandra Retico;Roberto Bellotti;Paolo Bosco;Piergiorgio Cerello;Gianfranco Gargano;Ernesto Lopez Torres;Rosario Megna;Marco Peccarisi;Maria Evelina Fantacci

Abstract: Purpose: The aim of this work is to evaluate the potential of combining different Computer Aided Detection (CADe) methods to increase the actual support for radiologists of automated systems in the identification of pulmonary nodules in CT scans.
Methods: The outputs of three different CADe systems developed by researchers of the Italian MAGIC-5 collaboration were combined. The systems are: the CAMCADe (based on a Channeler-Ant-Model which segments vessel tree and nodule candidates and a neural classifier), the RGVPCADe (a Region-Growing-Volume-Plateau algorithm detects nodule candidates and a neural network reduces false positives); the VBNACADe (two dedicated procedures, based respectively on a 3D dot-enhancement algorithm and on intersections of pleura surface normals, identifies internal and juxtapleural nodules, and a Voxel-Based-Neural-Approach reduces false positives. A dedicated OsiriX plugin implemented with the Cocoa environments of MacOSX allows annotating nodules and visualizing singles and combined CADe findings.
Results: The combined CADe has been tested on thin slice (lower than 2 mm) CTs of the LIDC public research database and the results have been compared with those obtained by the single systems. The FROC (Free Receiver Operating Characteristic) curves show better results than the best of the single approaches.
Conclusions: Has been demonstrated that the combination of different approaches offers better results than each single CADe system. A clinical validation of the combined CADe as second reader is being addressed by means of the dedicated OsiriX plugin.

authors' response to reviewers' comments

Click here to download authors' response to reviewers' comments: AnswerToRefereesJCARS2ndrevision.pdf

---
Reviewers' comments:

Reviewer #1:

My main concern-

1) Comparing with the references [10] [11], this paper does not add more significance and novelty in terms of methodology and results.

Re: In reference [10], the CAD algorithms were evaluated on an homogeneous dataset (in terms of acquisition and reconstruction parameters). Since no training set was available, each CAD system participating to the Anode09[10] competition was trained on a different dataset.
The main difference between [10] and this paper is that, in this work, the three CADs were trained on the same heterogeneous dataset of LIDC (i.e. different protocols, manufacturer of the machine, etc etc).
In principle, it would have been possible to attribute the gain in combining CAD systems only to differences in the train datasets.
This work is indeed showing that even with a common training dataset the improvement in combining different CADs is not negligible, due to the intrinsic differences in the CAD algorithms.

2) This paper lacks sufficient details of the evaluation and validation of the combined system. Without them, the results are not convincing and justified.

Re: We added more details, now the evaluation procedure should be more clear.

Introduction

Lung cancer is one of the most important health issues in developed countries and the most common cause of cancer-related deaths, with about 28% and 19% of all cancer-related deaths in the United States [1] and in European Union [2], respectively.

Lung cancer most commonly manifests itself as non-calcified pulmonary nodules. Computed Tomography (CT) was shown to be the most sensitive imaging modality for the detection of pulmonary nodules, particularly after the introduction of the multi-detector-row and helical CT technologies [3]. Therefore, screening programs based on low-dose CT are regarded as a promising approach for detecting early-stage lung cancers [4] and reducing the number of lung cancer deaths, as recently confirmed by the U.S. National Cancer Institute in its release of early results from the National Lung Screening Trial (NLST) [5]. The amount of data that need to be interpreted in screening CT examinations can be very large, especially when multi-detector helical CT and thin collimation are used, thus generating up to about 500 2-dimensional images per scan. It was indeed demonstrated that a large number of nodules (20–35%) can be missed in screening diagnoses [6].

In such a scenario, computer aided detection (CAD) methods could be very useful in supporting radiologists in the identification of early-stage pathological objects. It has already been demonstrated by several studies [7-9] that, in addition to a considerable time saving, the sensitivity of radiologists assisted by computer aided detection (CAD) systems improves with respect to the performance of the radiologists alone.

However, as it is shown in this paper and in [10,11], very good performances can be achieved combining different CAD systems, more than a developing a single optimal approach. In prior work [10,11] it was shown that there is always a combination of CAD that over-performed the best CAD scheme. It is likely that different methods have complementary strengths. Along these lines, the MAGIC-5 Italian Collaboration [12,13] adopted a development strategy that led to three different CADe algorithms for automated lung nodule identification. The present paper discusses how the three approaches were combined to generate a common system, the M5 CAD..

Methods and materials

The three CADe system prototypes, described more in detail in the following, are:

- the ^{CAM}CAD (Channeler Ant Model CAD), in which, after the lung volume identification [14], the Channeler Ant Model is used to segment the vessel tree and the nodule candidates and a Feed-Forward Neural Network (FFNN) is implemented in order to classify the segmented objects [15-17];
- the ^{RGVP}CAD (Region Growing Volume Plateau CAD), in which, after the lung volume identification [14], a region growing algorithm is iteratively applied to the lung volume to detect nodule candidates, a thresholding on the candidate volume and sphericity and a neural network classifier are applied to reject the false positive findings [17,18];
- the ^{VBNA}CAD (Voxel Based Neural Approach CAD), in which, after the lung volume segmentation, a 3D Dot-Enhancement (DE) algorithm identifies the internal nodule candidates [19,20] then, a procedure enhancing regions where many pleura surface normals intersect provides the juxtapleural nodule candidates [21] and a SVM (Support Vector Machine) classifier working at the voxel level reduces the amount of false positives [17, 22].

The public research database LIDC [23] was used to train and the validate all the CAD systems.

The dataset

The dataset used for this study consists of 138 CT scans from the LIDC [23] database, the biggest publicly available collection of annotated CTs. LIDC is a multi-centre and multi manufacturer database, currently under development, popolated with CTs of different collimation, kVp, tube current and reconstructed slice thickness. It therefore provides a general sample which is likely to realistically represent the input from a large scale multi-centre screening program. Presently, several hundred cases of the LIDC database are available; for this study 138 CT “thin slice” scans (i.e., scans with a slice thickness in the 0.5 to 2.0 mm range), were selected.

In order to capture the inter-reader variability the LIDC consortium provides four different annotations made by four expert radiologists for each case, obtained in a two-phase reading modality. The LIDC annotations contain three kinds of objects

[24]: nodules with diameters > 3mm, nodules with diameters < 3mm and "false positives" with diameters > 3mm. The contours of the objects marked as nodules with a diameter > 3mm were provided by every reader together with eight subjective characteristics in a 1 − 5 scale: subtlety, internal structure, calcification, sphericity, margin, spiculation, texture, malignancy. The CT scans used for the analysis were acquired at a 120 kVp voltage, with a current varying from 40 mA to 172 mA, a pixel spacing in the 0.434 mm to 0.762 mm range and a number of single slices ranging from 154 to 730.

The selected dataset was randomly divided in two homogeneous subsets, each containing 69 CTs, which were used for the training/optimization and for the validation procedures of the three CAD systems, respectively.

The Gold Standard reference was defined as the group of nodules with diameters > 3mm annotated by at least two radiologists. The nodules with diameters > 3mm annotated by only one radiologist were not considered as false findings in the evaluation of the FROC curves. A nodule has been visually labeled either as internal, if fully contained in the lung parenchyma, or as juxtapleural, if connected to the lung volume border. According to these criteria, the training dataset contains a total of 138 (96 internal and 42 juxtapleural) nodules, while the validation dataset contains a total of 114 (95 internal and 19 juxtapleural) nodules.

The Channeler Ant Model CAD

The full CT analysis is a sequence of four functional modules: lung segmentation, nodule hunter, filtering stage and neural network classification.

Lung segmentation

Lung segmentation aim is to extract the lung parenchyma from the whole CT scan. A brief description of the algorithm is:

- an optimal gray-tone threshold θ_0 for the segmentation of the respiratory apparatus is found by the analysis of Hounsfield histogram;
- a 3D region growing with simple threshold is applied to the whole CT volume; consequently, a voxel is included in the region to be grown, if its Hounsfield value is smaller than θ_0. The output of this procedure is a binary mask of the respiratory system containing trachea, bronchi and lungs;

- the correct handling of the hilar region is obtained by removing external airways through a wave-front simulation model with a proper stop condition. The output of this procedure is a mask containing only the lungs;
- incidental fusion of the lungs is corrected using a new threshold value θ_0, applied only to the fusion zone;
- a morphological 3D closing includes pleural and internal nodules and patches the vessel concavities.

This lung-segmentation module is also used in the ^{RGVP}CAD system.

Nodule Hunting

The Channeler Ant Model [15,16] is used as a segmentation method for the vessel tree and the nodule candidates. The approach is iterative, it consists in a sequence of two independent deployments for the right and left lungs.

The first ant colony segments the vessel tree, starting from an anthill in the vicinity of the root of the tree. Ants explore (i.e. live in) a 3D environment described in terms of positions and intensities of voxels. Their life cycle is a sequence of atomic time steps, during which they behave according to a predefined set of rules [15]: they release pheromone while moving in the 3D environment defined by the lung volume; they also change their energy, so as to be able to reproduce or die depending on its value. The environment is defined by the voxel image intensities, which can be thought of as the amount of available food for the colony: therefore, voxel intensities should be progressively consumed when the number of visits increases. This mechanism, required to make the colony evolve and explore the environment, is implemented in a complementary way: whenever the limit to the maximum number of visits in a voxel is reached, the voxel is no more available as a destination. When all the ants in the colony have died, the process stops, the segmented object is removed from the original image and its coordinates are added to a list.

In the remaining image, any voxel with intensity above a predefined threshold is a new anthill and a new ant colony is deployed. If the size of an object is large with respect to the maximum expected size of a nodule, as it happens with the bronchial and vascular trees, the object is processed and smaller connected objects are looked for. The procedure is repeated by trying as anthill each voxel in the

lung volume with an intensity larger than -700 Hounsfield Units: when no more voxels meet the conditions to become anthills, the information provided by the pheromone map is analyzed.

An iterative analysis is carried out: each voxel with a pheromone content above the minimum accepted value (8,000 units) is used as a seed for a region growing with an adaptive threshold. The threshold value is lowered iteratively for each seed and the selected value is the one corresponding to the minimum growth of the region when the hypothetical threshold is lowered by a quantum.

Whenever a region is larger than a preset value (50 voxels), it is further analyzed in search of nodule candidates connected to it. In order to do so, a rolling sphere scans the finding and disentangles spherical-like structures. The procedure is repeated three times, with spheres of increasing initial radius (1.5, 2.5, 3.5 mm).

In short, a full sequence of ant colony deployments generates a pheromone release map that is analyzed by a dedicated filter, which turns it into a list of candidate findings, each defined by a list of voxels and the values of a set of features related to their geometrical properties, their intensity pattern, their location in the lung.

Filtering

The number of candidates per CT, although depending on the number of slices, ranges between several hundreds to a few thousands per scan, a number far too large to be used as input for a neural network classifier. However, the vast majority of findings is easily rejected with some selections that make use of the correlation between few of the evaluated features: the radius, the sphericity, the fraction of voxels connected to the cage, the so called attach flag (AF), which identifies whether the finding is isolated (AF = 0) or not (AF > 0). If the finding is attached to a larger structure (i.e., the vessel tree), AF is related to the size of the rolling sphere and can range from 1 to 3.

The filtering is performed with a cut function on the histogram that correlates the sphericity to the radius: findings with a sphericity below the cut value at any given radius are rejected. Given the different way in which findings are extracted from the pheromone map, for each AF value the function parameters are different. The filtering level is defined as a compromise between the requirement of maintaining a high sensitivity and the goal of forwarding as less as possible findings to the classification stage.

An additional filter requires the fraction of voxels connected to the cage to be smaller than 0.6, in order to get rid of elongated artifacts attached to the cage.
After the filtering stage, the average number of findings in the training/testing and validation set is 27 and 25, respectively.
For the training/testing set, the sensitivity and the number of false positives/scan, defining the first point on the FROC curve obtained with the classification module, are 0.81 and 0.83, respectively.

Classification

Until the end of the filtering stage, very few of the nodule candidate features are used. In particular, no direct information about the image intensities in the candidate voxels is taken into account. The selected set of features for the classification stage makes use of properties that describe the finding size, shape, location, intensity (inside and on the border), as well as the above-defined AF value, which corresponds to different parameters of the nodule hunting algorithm.
The classification was carried out with a four layer FFNN: 11 neurons in the input layer, 25 and 7 in the intermediate layers and one in the output layer.
The full list of features for the input layer follows: sphericity, radius, Shannon entropy of the inner and the border voxels, skewness, kurtosis, average and standard deviation of the voxel intensities, fraction of voxels connected to the cage, AF value. The classification was optimized on the training/testing sample of 69 CTs and 138 true findings, with a cross validation procedure: 30 sub-list of true findings and false findings were classified as testing sample against all the other true and false findings used as training sample.

The Region Growing Volume Plateau CAD

The ^{RGVP}CAD system is an upgrade of the CAD system presented in a previous paper [18]. At present the ^{RGVP}CAD system consists of four main modules:

- the lung segmentation, previously described in the lung segmentation section of the ^{CAM}CAD;
- nodule candidate detection performed through an iterative region growing;
- statistical and morphological feature extraction;
- candidate nodule classification by means of linear filters and a FFNN.

Nodule Hunting

The nodule candidate detection is carried out through an iterative region growing. The RG inclusion rule is the logical AND of the following two rules (fig. 1):

1) Simple Bottom Threshold (SBT) rule, i.e. a voxel is included in the region if its Hounsfield value is greater than a threshold θ_1;
2) Mean Bottom Threshold (MBT) rule, i.e. a voxel is included in the region if the average of the Hounsfield values of the voxel and its 26 neighbors is greater than a threshold θ_2.

The θ_1 threshold is fixed, while the θ_2 threshold is varied in the range $[\theta_b, \theta_t]$ until an optimal value is found for every nodule candidate, on the basis of size-related properties: a graph of the region volume (or number of voxels) as a function of the θ_2 value is generated (fig. 2) and θ_2 is set to a value corresponding to the plateau. Using this procedure with a value of θ_b close to the Hounsfield unit of air and appropriate values for θ_1 and θt, it is possible to obtain an optimal segmentation of the nodule candidates in the lung parenchyma, without any prior knowledge of their average Hounsfield value.

Filtering

A significant number of statistical and morphological features is calculated in the third module of the ^{RGVP}CAD system. The initial set of features for each candidate nodule is 17: Volume, Radius Variance, Over Radius Variance, Radius Standard Deviation, Over Radius Standard Deviation, Radius Skewness, Radius Kurtosis, Over Radius Kurtosis, Radius Compactness (radius - error radius ratio), Over Radius Compactness, Maximum Distance (from center to farthest voxel), Roundness, Ellipticity, Maximum Hounsfield intensity, Standard deviation of Hounsfield intensity, Shannon's Entropy of Hounsfield intensity distribution. The *Over* prefix refers to those features that are evaluated only for voxels whose distance from the the candidate center of mass is greater than the average radius of the candidate nodule. The best discriminant features, selected by means of the Sequential Backward Selection (SBS) procedure [25], are: Volume, Over Radius Kurtosis, Radius Compactness, Maximum Distance, Maximum Hounsfield intensity, Standard deviation of Hounsfield intensity, Shannon's Entropy of Hounsfield intensity distribution.

Classification

The candidate nodule classification consists of two linear filters on the candidate features, volume and roundness, a FFNN is then applied. The filtering is applied to reduce the great number of false positive candidates generated by the second module of the CAD system: only candidate nodules with both volume and roundness above a minimum value are selected as candidate nodules for the next step. After this filtering, the candidate nodules are classified by means a FFNN, with 7 neurons in the input layer, 3 in the hidden layer and one in the output layer. The training procedure applied was the back-propagation algorithm. In order to optimize the parameters and to train the ^{RGVP}CAD system a 2-fold cross-validation technique [26] was used.

The Voxel Based Neural Approach CAD

The ^{VBNA}CAD system deals differently with internal and juxtapleural nodules, by means of two dedicated procedures: CAD_I for internal and CAD_{JP} for juxtapleural nodules. Both are three-step procedures [17, 19-22]. The first step consists in the lung segmentation; the second step consists in the ROI (Region Of Interest) hunter and performs the candidate nodule selection; the third step consists in the FP reduction. For the last step, an original procedure, the Voxel-Based Neural Approach is implemented to reduce the number of FPs in the lists of internal and juxtapleural candidate nodules.

Lung segmentation

A three-steps approach based on thresholding, region growing and morphological operators is implemented [27]:

• once the scans have been isotropically resampled, to separate the low-intensity lung parenchyma from the high-intensity surrounding tissue (fat tissue and bones), the voxel intensities are thresholded at a fixed value;

• in order to discard all the regions not belonging to the lungs, the biggest connected component not touching the boundary of the volume is considered;

• vessels and airways are not in included in the segmented lung at this stage since their volume is outside the segmented lung volume. To include them without modifying the pleura surface morphology, i.e. without modifying the shape of

pleura irregularities (including juxtapleural nodules), a combination of morphological operators is applied. In particular, a sequence of the dilation and the erosion operators with spherical kernels r_d and r_e, with $r_e > r_d$ is implemented.
Finally, the logical OR operation between the so-obtained mask and the original lung mask provides the final mask, where the vessels and the airway walls are filled in, while maintaining the original shape of the lung border.
The identified lung mask is used for CAD_I, whereas its boundary is used for CAD_{JP}.

ROI hunter for internal nodules

In the CADI the internal nodules are modelled as spherical objects with a Gaussian profile, following the approach proposed in [28]. To detect this kind of objects, a dedicated DE filter is implemented. The filter determines the local geometrical characteristics of each voxel by evaluating the function z_{dot} using the eigenvalues of the Hessian matrix:

$$z_{\mathrm{dot}}(e_1, e_2, e_3) = \begin{cases} |e_3|^2/|e_1| & \text{if } e_1<0\, e_2<0\, e_3<0 \\ 0 & \text{otherwise} \end{cases}$$

where e_1, e_2, e_3 are the eigenvalues of the Hessian matrix for each voxel, sorted so that $|e_1|>|e_2|>|e_3|$ |
To enhance the sensitivity of this filter to nodules of different sizes, a multi-scale approach has to be followed. This approach, combines the z_{dot} function with Gaussian smoothing at several scales with the prescriptions given in [28]. The result of the filter is a matrix

$$z_{max} = max(s_i^2 z_{\mathrm{dot}}(s_i)) \text{ with i=1,..,N}$$

where s_i are the sigma of the Gaussian smoothing.
Local maxima of the matrix filtered by the dot-enahncement are the internal candidate nodule locations. A large number of false positives is included at this stage, above all crossings between blood vessels.

ROI hunter for juxtapleural nodules

In the CADJP, in order to identify juxtapleural candidate nodules, pleura surface normals are constructed and each voxel is assigned a score proportional to the number of normals intersecting in it. Normals are evaluated using the triangular mesh representing the pleura surface, obtained applying the marching cube

algorithm on the lung mask. In particular, the normal to each triangle is calculated by using the vector product between the triangle edges; then the normals to each mesh vertex are evaluated averaging all the triangle normals of the neighboring triangles.

Since the evaluation of the normal intersections in the real 3D space is a complex and computationally intensive operation, it is implemented in the voxel space. This means that each voxel is associated a score proportional to the number of normals passing through it. To deal with noise, cylinders with Gaussian profile are considered instead of segments [21,29]. This information is collected in the score matrix S(x,y,z). The local maxima of the 3D matrix S(x,y,z) are the juxtapleural candidate nodule locations. Also in this case a large number of FPs is found, mostly due to irregularities in the pleura surface (e.g. apical scars, pleural thickening and plaques) and movement artifacts.

Classification

In order to classify the candidate nodule findings obtained in the previous step, an original procedure, the Voxel-Based Neural Approach [20-22], performs the reduction of the number of FPs in the lists of internal and juxtapleural candidate nodules. First, a ROI including voxels belonging to the candidate nodule is defined from each location provided by the previous step. The basic idea of the VBNA is to associate to each voxel of a ROI a feature vector defined by the intensity values of its 3D neighbours (in this case 5x5x5 intensity values) and the eigenvalues of the gradient matrix and of the Hessian matrix. In the original VBNA method the feature vectors were then classified by a three-layer FFNN which was trained to assign each voxel either to the nodule or normal tissue target class. In this paper a different classifier is implemented at this stage: Support Vector Machines (SVM) are used instead of neural networks. At the end of the procedure, each ROI is assigned a degree of suspicion averaging the score of all the voxels belonging to it.

CAD_I and CAD_{JP} combination

In principle, CAD_I and CAD_{JP} act on two complementary regions of the image. In practice, due to imperfections in the lung segmentation, i.e. pleural nodules under-segmented by the procedure, it may happen that the same object is detected

by the two procedures. To prevent having two marks pointing the same object, a simple clustering rule is applied: if the distance between the centres two findings is less than 4.0 mm, the finding with lower probability is discarded. This clustering rule is based on the fact this kind of collisions happen mostly for nodule objects.

The combination procedure

As shown in [10,11], there is no reason to assume that a single CAD scheme would be optimal for nodule detection. It is more likely that different methods have complementary strengths, as already shown on different datasets in [10] and [11]. The outputs of the three CAD sub-systems described are evaluated and combined following the same procedure adopted for the ANODE09 study [10,11]. The resulting CAD system is referred to as M5 CAD system. The findings of each CAD sub-system must be considered in terms of their degree of suspicion (i.e. likelihood or probability to be a true nodule) p, which is the final output of the procedure of candidate nodules classification for the three separate sub-systems.

In order to combine findings from different CAD sub-systems, a normalization of the finding probabilities is needed [11]. This operation is carried out by associating a new value $f(p)$ to each finding with degree of suspicion p.

The new degree of suspicion $f(p)$ is evaluated according to the performance obtained by the corresponding CAD system on the validation set, i.e. evaluating for each finding with probability p

$$f(p) = \frac{TP}{FP + TP + 1},$$

where TP (FP) is the number of true (false) positives obtained by considering all the CAD findings with $p_i >= p$. Of course this procedure requires to know the annotations and the performance of each CAD system on a selected set of data.

The $f(p)$ values can therefore be considered as the probability that a finding in the validation set with likelihood p or higher represents a true nodule.

The function $f(p)$ is computed for every finding from every sub-system. All findings are then sorted so as f_i, i = 1… n and $f_i >= f_j$ if $i < j$. Starting at f_i with i = 1, all findings f_j , $j = i + 1$… n are checked against a “matching condition” defined

by a preselected clustering distance. If two findings f_i and f_j match, then $f_i = f_i + f_j$, and f_j is removed from the list of findings.

Results

The results obtained for the separate systems on the 69 CTs of the validation dataset, containing a total of 114 nodules annotated by at least two radiologists, are reported in terms of FROC (Free-response Receiver Operating Characteristic) curves in Fig. 3. The FROC curves are evaluated using the following matching criterion [10]: a CAD finding is considered a true positive if its Euclidean distance from the center of the lesion annotated by the radiologists is less than m=1.5 times the radius of the annotated lesion.

Among the presented CAD systems, the best performance is provided by the CAM algorithm, which starts the classification step at about 85% sensitivity and 25 FP findings/CT exam. The VBNA and RGVP algorithms are 2 and 4 times less selective in the nodule hunting stage, respectively.

In addition to the FROC, the CAD performance was also evaluated by means of a Score Value SV, as defined in [10]: the average sensitivity at 7 fixed thresholds (1/8,1/4,1/2,1,2,4,8 FP/CT) along the FROC curve:

$$SV = (Sens_{1/8} + Sens_{1/4} + Sens_{1/2} + Sens_{1} + Sens_{2} + Sens_{4} + Sens_{8}) / 7$$

In order to verify that the three CAD systems are not too sensitive to the FROC matching criterion, the score SV was evaluated varying the *m* factor [10]: the results reported in Fig. 4 show that indeed the CAD systems performances are almost independent of *m* and that the selected matching condition (m = 1.5) is a stable choice.

The algorithm combination, computed considering a clustering distance of 3 mm and reported in Fig. 5 for the 0-10 false positive findings/CT range, confirms that there is an added value in combining different nodule hunting approaches.

In order to quantify the improvement introduced by the combination, the Score Values SV were computed for m=1.5. The results, reported in tab.1, show that the combined M5 CAD performs better than the CAM CAD of about 5% on average in the 1/8 to 8 FP/CT range.

	CAM	VBNA	RGVP	Combination
Score Value (SV)	0.57	0.43	0.39	0.62

Table 1: Score Value (SV) for the separate CAD-subsystems and for the combined M5 CAD.

The number of findings from different systems that cluster during the combination procedure provides a qualitative explanation of the reason why the combination procedure is helpful: the number of detected true nodules increases with the number of CADs in the combination, whereas the number of FP decreases (see tab.2).

Moreover, it is possible to evaluate the average probability increase of each clustered finding, defined as the difference between its probability f(p) after the combination and the highest probability among those of the findings that contribute to the combination. For TPs and FPs the average probability increase is 0.67 and 0.04, respectively.

The previous analysis shows, once more and on a different dataset, that combining different CAD system is helpful mostly because of the complementarity of the FP findings.

	1 CAD	2 CAD	3 CAD
Number of TP	18	25	65
Number of FP	36566	1410	139

Table 2: Number of true and false findings detected by 1, 2 and 3 CAD-subsystems. While the number of true findings increases when considering more than one sub-system, the opposite happens for the false findings.

As shown in fig.5, the overall sensitivity improvement in the combination is minimal at high values of FP findings/CT exam; on the other hand, the number of FP findings/CT exam at which a given sensitivity is reached is much smaller (about one half) when the three algorithms are combined. In other words, the time spent by the radiologists in reviewing and rejecting false CAD findings would be much smaller with a combined system.

Discussion

The advantages of a multi-thread approach in the automated search for nodules in lung CTs were already addressed in [10,11], where algorithms trained on different datasets (reviewed with different annotation protocols), proved to be complementary either in finding true nodules or in rejecting false findings.

The present analysis confirms that a multi-thread approach provides better results than a stand-alone one, even when the algorithms are trained on the same dataset, although in this case the most important advantage lies in the rejection of false positive findings.

In order to confirm that the M5 combined CAD system is actually useful to improve the performance of human experts, a clinical validation that will measure the contribution to the radiologists annotation of the three CAD sub-systems and their combination by defining a second reader protocol is being started. For this purpose a new dedicated OsiriX [30] plugin was developed (see fig.6). The plugin allows nodule annotation and visualisation: findings are shown in different colours for the separate CAD sub-system results, their combination and the radiologist annotation. It is also possible to select, using the sliders, different working points for each CAD and visualise only findings above the selected threshold. An example of the visualisation of a nodule detected by the three CAD sub-systems is shown in fig.7.

Conclusions

In this study, the performance of three different CAD sub-systems developed by the Italian MAGIC-5 Collaboration are presented. The algorithms, based on different approaches, were applied using the same working conditions on datasets from the LIDC database and the algorithms' performances were evaluated separately and combined. In a previous work [10] it was demonstrated that different systems compared on the same database show different results, due both to the underlying algorithm and the dataset that is used to train the classifiers or to set the internal model parameters, and that the combination of systems may yield relevant improvements due to the different ability in detecting different categories of nodules.

In the present work we demonstrate that even under the same operating conditions different CAD sub-systems have varying strengths and weaknesses. The result of combining them reveals how complementary they are, especially in the range of false positive findings up to 10/CT exam, the most interesting for radiologists. The overall performance of the M5 CAD is very good, with a sensitivity that reaches 80% at about 3 FP/scan, even with a loose selection of the Gold Standard reference sample (at least two radiologists out of four that annotated the CTs).

References

[1] Online document. American Cancer Society 2009, www.cancer.org/Research/CancerFactsFigures Accessed 16 December 2010.

[2] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008. Int J Cancer 117(12) 2893-2917.

[3] Diederich S, Lentschig MG, Overbeck TR, Wormanns D and Heindel W. 2001. Detection of pulmonary nodules at spiral CT: comparison of maximum intensity projection sliding slabs and single-image reporting. European Radiology, 11(8):1345–1350.

[4] The International Early Lung Cancer Action Program Investigators, 2006. Survival of Patients with Stage I Lung Cancer Detected on CT Screening. New England Journal of Medicine 355 (17), 1763-1771.

[5] Online document U.S. National Cancer Institute November 2010. http://www.cancer.gov/clinicaltrials/noteworthy-trials/nlst Accessed 16 December 2010.

[6] Roberts HC, Patsios D, Kucharczyk DM, Paul N, and Roberts TP, 2005, The utility of computer-aided detection (CAD) for lung cancer screening using low-dose CT, Computer Assisted Radiology and Surgery, Proceedings of the 19th International Congress and Exhibition, Berlin, June 22 - 25, 2005, International Congress Series 1281, pp. 1137–1142.

[7] Das, M. et al., 2006. Small Pulmonary Nodules: Effect of Two Computer-aided Detection Systems on Radiologist Performance. Radiology 241, 564-571.

[8] Brochu, B. et al., 2007. Computer-aided detection of lung nodules on thin collimation MDCT: impact on radiologists' performance. Journal de Radiologie, 88 (4), 573-578.

[9] Matsumoto, S. et al., 2008. Computer-aided detection of lung nodules on multidetector row computed tomography using three-dimensional analysis of nodule candidates and their surroundings. Radiation Medicine, 26 (9), 562-569.

[10] van Ginneken B, Armato SG 3rd, de Hoop B, van Amelsvoort-van de Vorst S, Duindam T, Niemeijer M, Murphy K, Schilham A, Retico A, Fantacci ME, Camarlinghi N, Bagagli F, Gori I, Hara T, Fujita H, Gargano G, Bellotti R, Tangaro S, Bolaños L, De Carlo F, Cerello P, Cristian Cheran S, Lopez Torres E, Prokop M., Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study., Med Image Anal. 2010 Dec;14(6):707-722.

[11] Niemeijer, M.; Loog, M.; Abràmoff, M.D.; Viergever, M.A.; Prokop, M.; van Ginneken, B.; , "On Combining Computer-Aided Detection Systems," Medical Imaging, IEEE Transactions on, vol.30, no.2, pp.215-223, Feb. 2011.

[12] Bellotti, R. et al., 2007. Distributed medical images analysis on a GRID infrastructure. Future Generation Computer Systems, 23, 475-484.

[13] Golosio, B. et al., 2009. A novel multithreshold method for nodule detection in lung CT. Medical Physics 36 (8), 3607-3618.

[14] De Nunzio, G. et al., 2009. Automatic Lung Segmentation in CT Images with Accurate Handling of the Hilar Region. Journal of Digital Imaging, Online First.

[15] Cerello P, Cheran SC, Bagagli F, Bagnasco S, Bellotti R, Bolanos L, Catanzariti E, De Nunzio G, Fiorina E, Gargano G, Gemme G, Lopez Torres E, Masala G, Peroni C and Santoro M. 2008. The Channeler Ant Model: Object segmentation with virtual ant colonies. In IEEE Nuclear Science Symposium Conference Records, 2008, 3147-3152.

[16] Cerello, P. et al., 2010. 3-D object segmentation using ant colonies. Pattern Recognition 43 (4), 1476-1490.

[17] Fantacci ME et al., Computer aided detection of nodules in low dose and thin slice lung CT, DOI: 10.1594/ecr2010/C-1053.

[18] Bellotti, R. et al., 2007. A CAD system for nodule detection in low-dose lung CTs based on Region Growing and a new Active Contour Model. Medical Physics, 34 (12), 4901-4910.

[19] Retico A, Delogu P, Fantacci ME, Gori I, and Preite Martinez A. 2008. Lung nodule detection in low-dose and thin-slice computed tomography. Comput. Biol. Med., 38(4):525–534.

[20] Gori I, Bagagli F, Fantacci ME, Preite Martinez A, Retico A, De Mitri I, Donadio S, Fulcheri C, Gargano G, Magro R, Santoro M and Stumbo S. 2007. Multi-scale analysis of lung computed tomography images. Journal of Instrumentation, 2(09):P09007.

[21] Retico A, Fantacci ME, Gori I, Kasae P, Golosio B, Piccioli A, Cerello P, De Nunzio G, Tangaro S 2009 Pleural nodule identification il low-dose and thin-slice lung computed tomography. Comput Biol Med. 39(12):1137-1144

[22] Retico, A. et al., 2009. A voxel-based neural approach (VBNA) to identify lung nodules in the ANODE09 study. Proceedings of SPIE, Vol. 7260, 72601S-72601S-8.

[23] McNitt-Gray MF, Armato SG 3rd, Meyer CR, Reeves AP, McLennan G, Pais RC, Freymann J, Brown MS, Engelmann RM, Bland PH, Laderach GE, Piker C, Guo J, Towfic Z, Qing DP, Yankelevitz DF, Aberle DR, van Beek EJ, MacMahon H, Kazerooni EA, Croft BY, Clarke LP. 2007. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. Acad. Radiol. 14(12), 1464-1474.

[24] https://wiki.nci.nih.gov/display/cip/lidc Accessed 16 december 2010

[25] Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. New York Academic Press.

[26] Stone M, 1974, Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society B 36 (1), 111-147.

[27] Retico, A., Bagagli, F., Camarlinghi, N., Carpentieri, C., Fantacci, M. E., and Gori, I., "A voxel-based neural approach (VBNA) to identify lung nodules in the ANODE09 study," in [Medical Imaging 2009: Computer-Aided Diagnosis], 7260, 72601S–8, SPIE, Lake Buena Vista, FL, USA (Feb. 2009).

[28] Li Q, Sone S, and Doi K. 2003. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. Medical Physics, 30(8):2040-2051.

[29] Paik SD, Beaulieu CF, Rubin GD, Acar B, Jeffrey RB Jr., Yee J, Dey J, and Napel S. 2004. Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. IEEE Trans Med Imaging, 23(6):661-675.

[30] http://www.osirix-viewer.com/ Accessed 16 december 2010

Fig.1 Illustration of the inclusion rule of the ^{RGVP}CAD system

Fig.2 A graph of the volume (expressed in terms of voxels) as a function of the threshold θ2 for a nodule candidate.The red circle indicates the selected θ_2 value.

Fig.3 FROC curves obtained by the three CAD systems, in the range 0-100 FP/CT: in red VBNA CAD is shown, in green RGVP CAD and in blue CAM CAD.

Fig.4 Sensitivity score value (SV) in the range of 1/8 to 8 FP/CT for the 3 CAD system as a function of the matching distance *m*.

Fig.5 The FROC curve obtained with the CAD combination compared with those obtained with the single systems in the range of FPs of radiological interest (0-10 FP/CT): the VBNA CAD, the RGVP CAD and the CAM CAD are shown in red, green and blue, respectively; the combination is shown in purple.

Fig.6 Example of a screenshot of the OsiriX plugin

Fig.7 An example of nodule visualisation

Fig 1

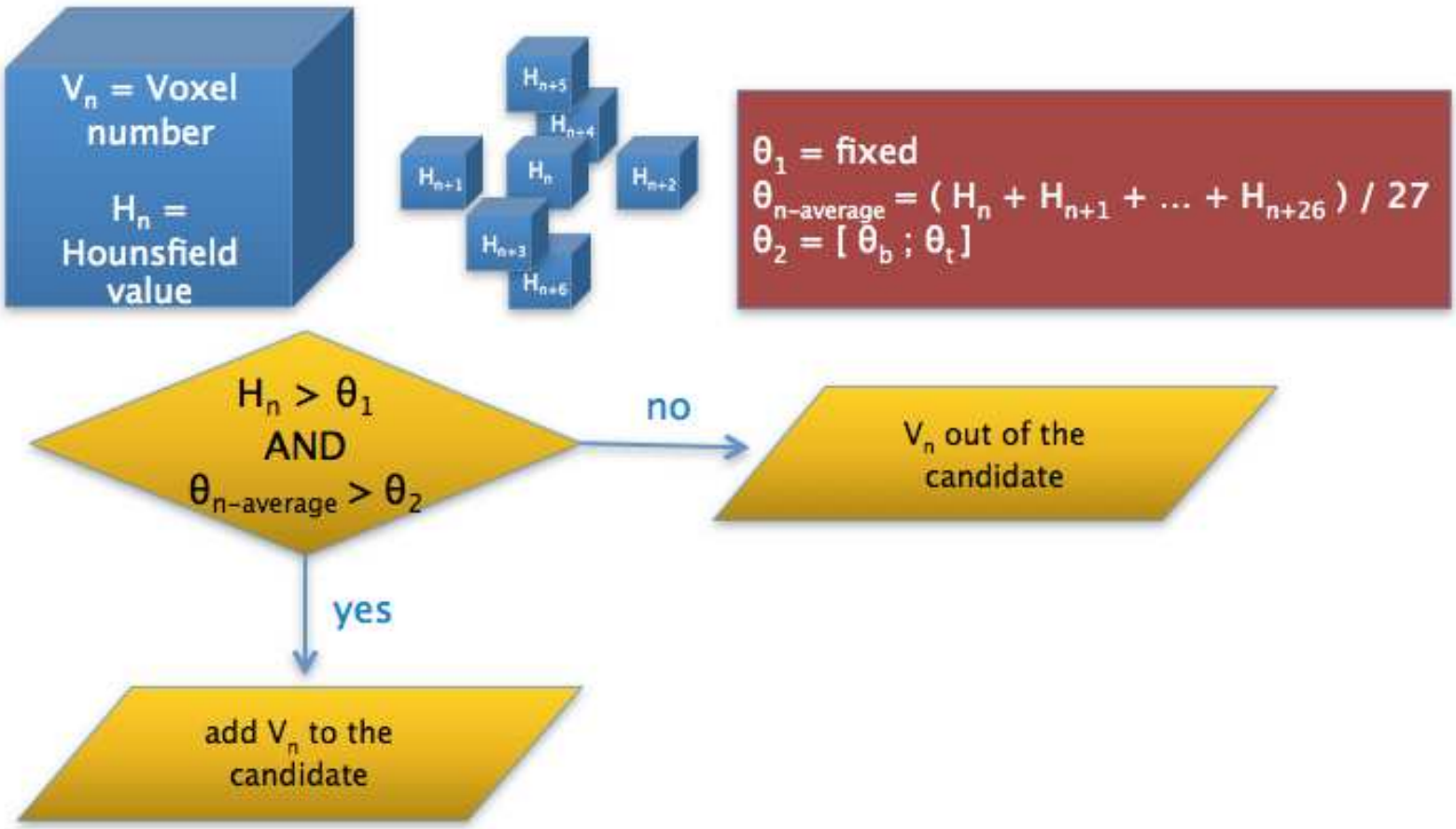
V_n = Voxel number
H_n = Hounsfield value
H_{n+5}
H_{n+4}
H_{n+1}
H_n
H_{n+2}
H_{n+3}
H_{n+6}
θ_1 = fixed
$\theta_{n\text{-average}} = (H_n + H_{n+1} + \dots + H_{n+26}) / 27$
$\theta_2 = [\theta_b ; \theta_t]$
$H_n > \theta_1$ AND $\theta_{n\text{-average}} > \theta_2$
no
V_n out of the candidate
yes
add V_n to the candidate

Fig 2

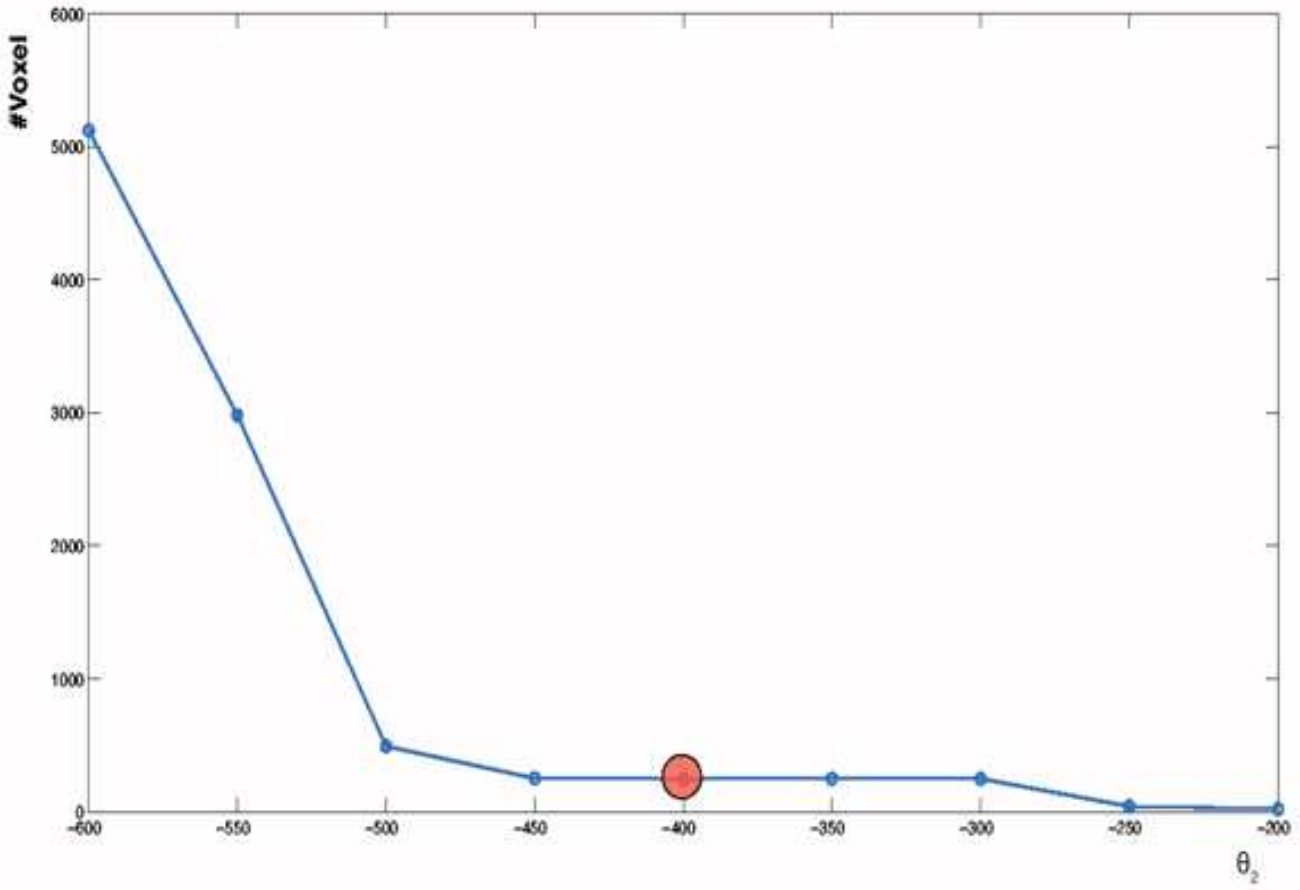
#Voxel
6000
5000
4000
3000
2000
1000
0
-600
-550
-500
-450
-400
-350
-300
-250
-200
θ2

Fig 3

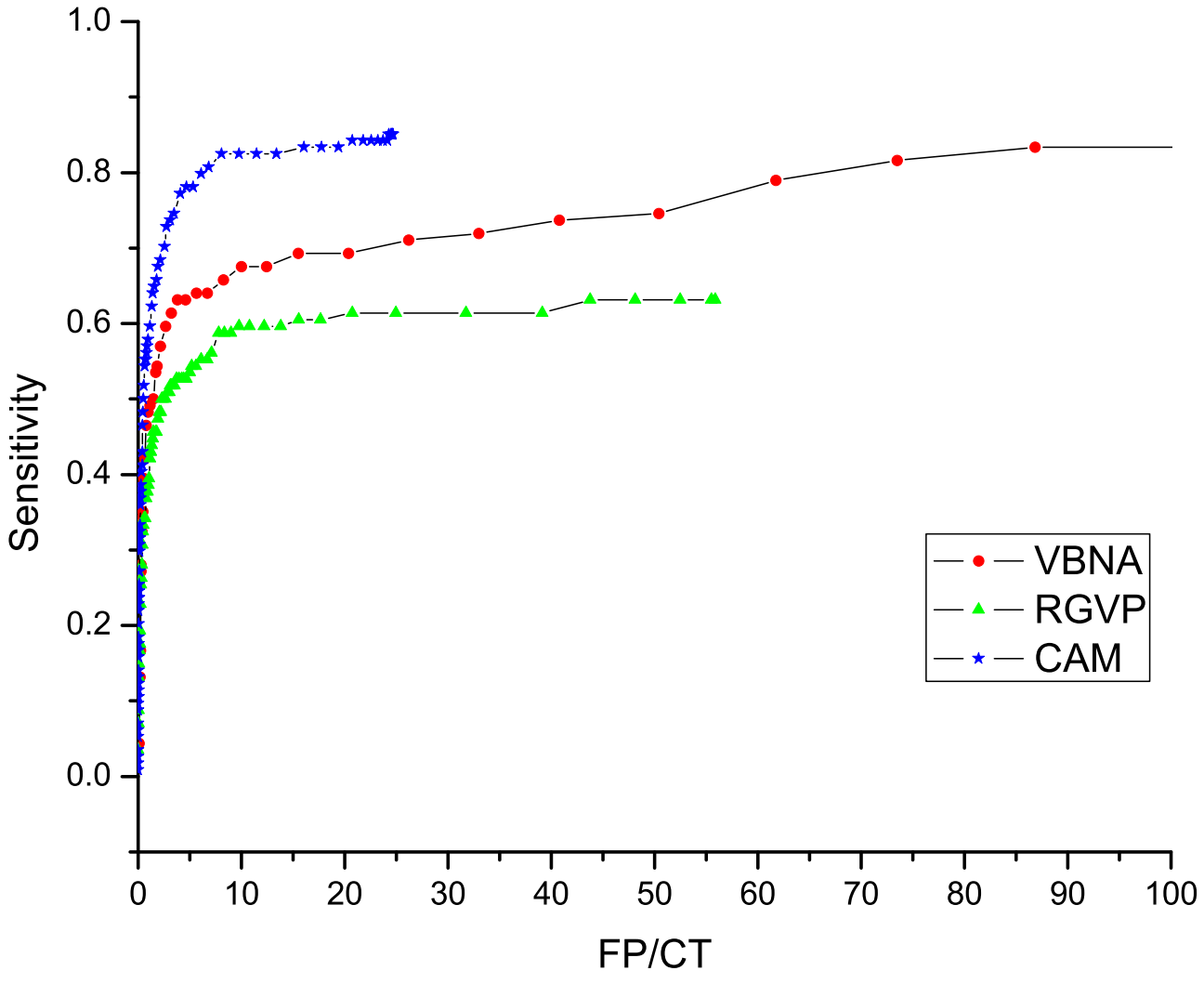
1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
0
10
20
30
40
50
60
70
80
90
100
FP/CT
VBNA
RGVP
CAM

Fig 4

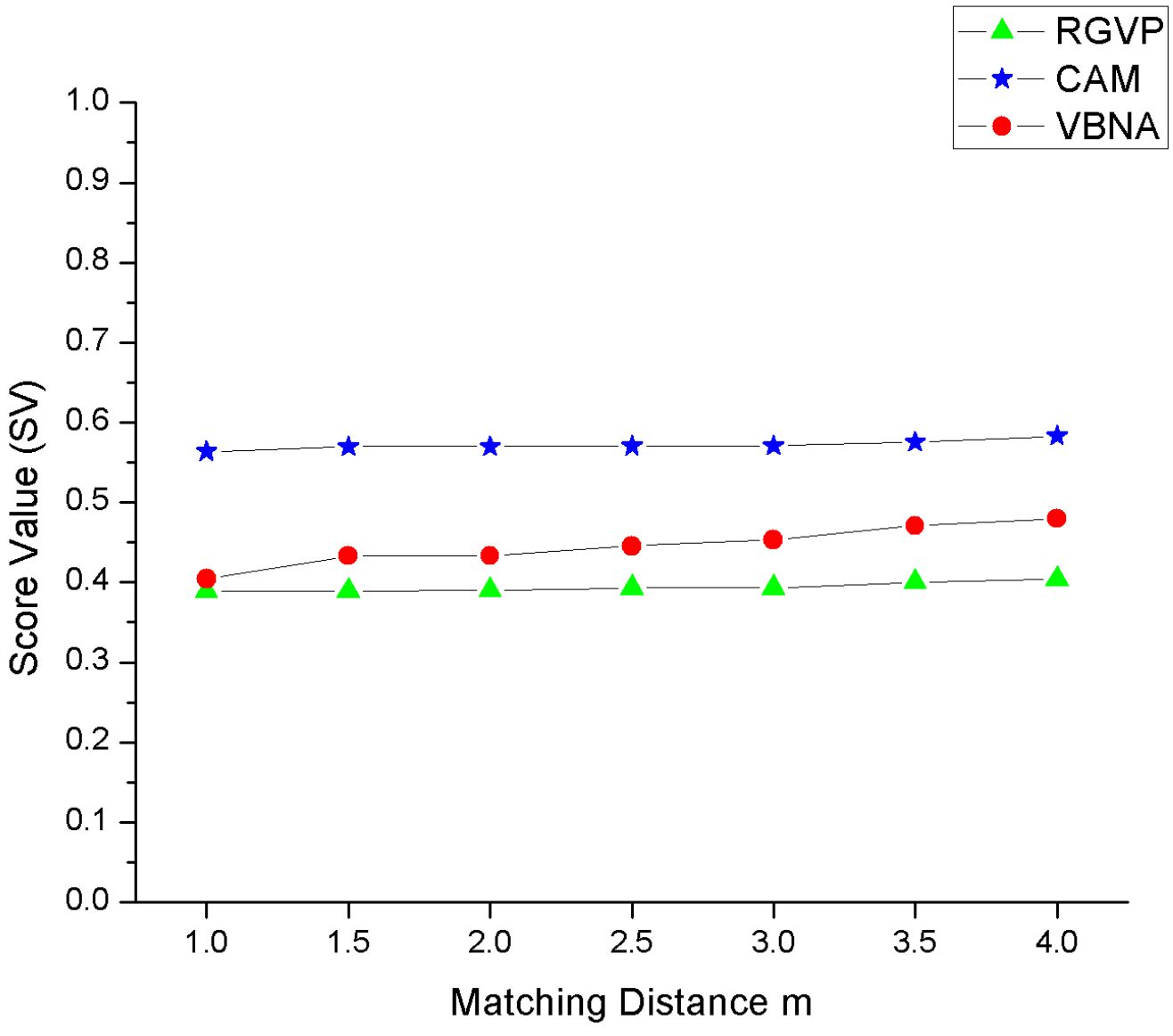
RGVP
CAM
VBNA
Score Value (SV)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Matching Distance m

Fig 5

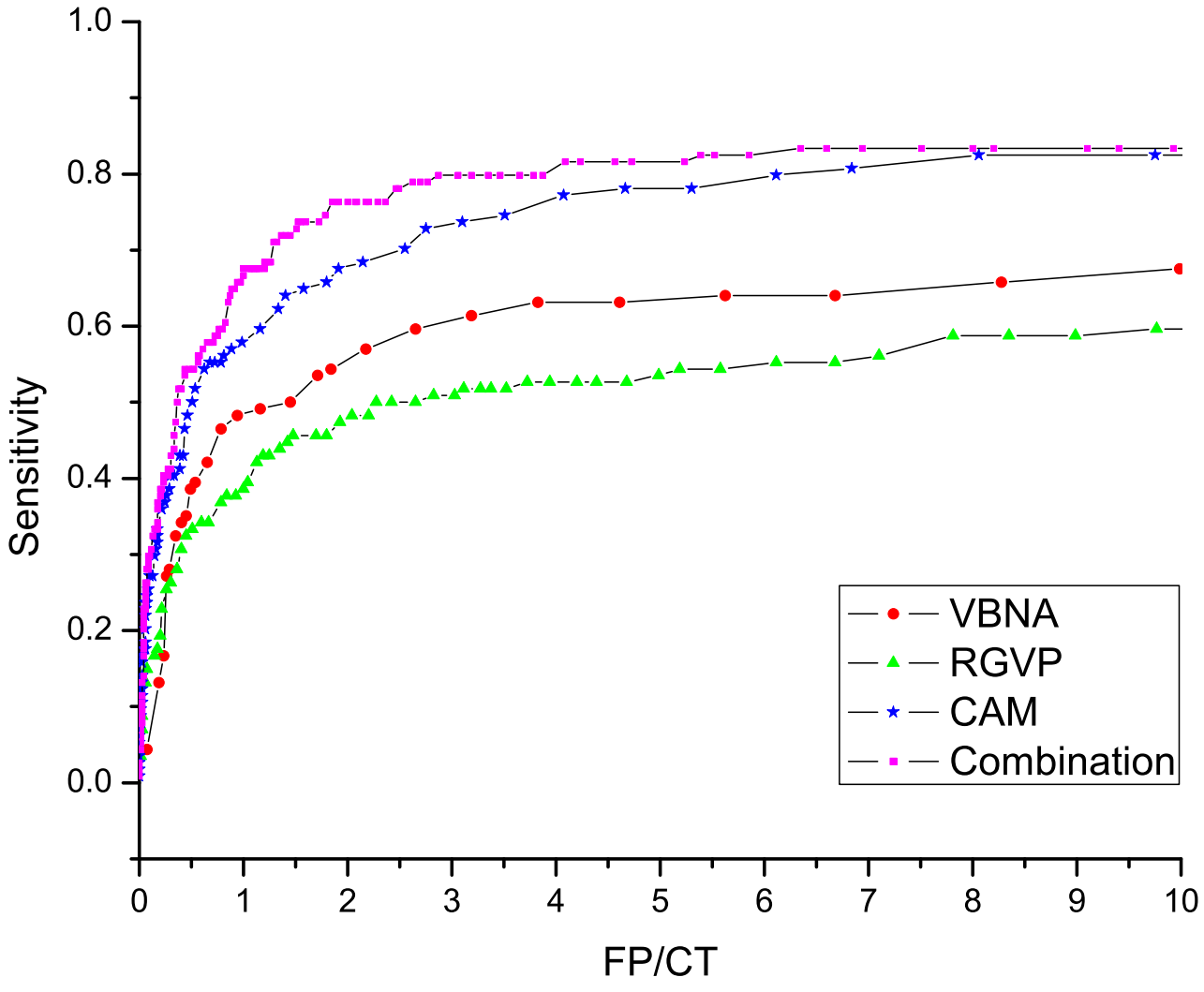
Sensitivity
FP/CT
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
VBNA
RGVP
CAM
Combination

Fig 6

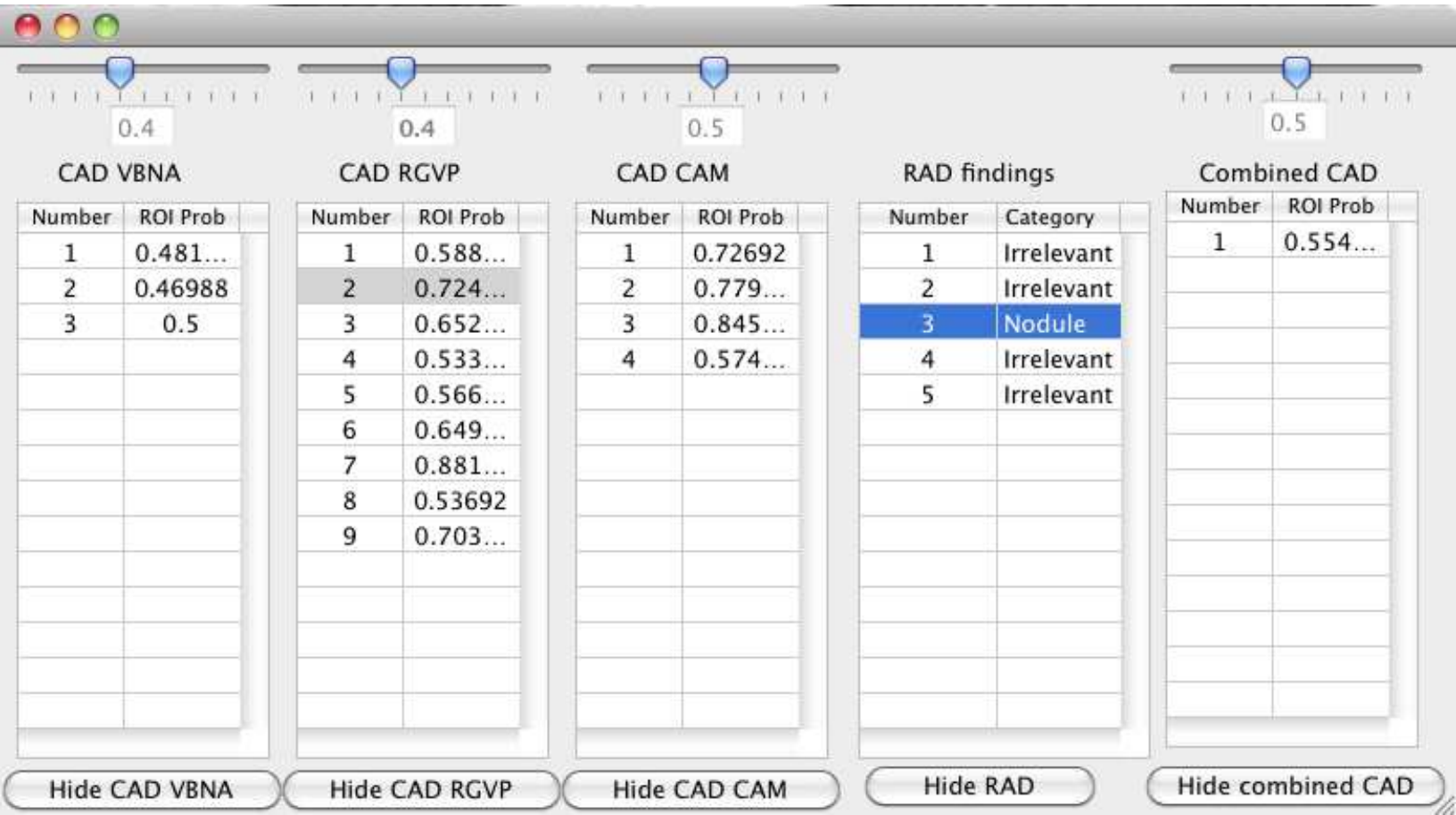
0.4
0.4
0.5
0.5
CAD VBNA
CAD RGVP
CAD CAM
RAD findings
Combined CAD
Number
ROI Prob
1 0.481...
2 0.46988
3 0.5
Number
ROI Prob
1 0.588...
2 0.724...
3 0.652...
4 0.533...
5 0.566...
6 0.649...
7 0.881...
8 0.53692
9 0.703...
Number
ROI Prob
1 0.72692
2 0.779...
3 0.845...
4 0.574...
Number
Category
1 Irrelevant
2 Irrelevant
3 Nodule
4 Irrelevant
5 Irrelevant
Number
ROI Prob
1 0.554...
Hide CAD VBNA
Hide CAD RGVP
Hide CAD CAM
Hide RAD
Hide combined CAD

Fig 7