

Automated multistucture atlas-assisted detection of lymph nodes using pelvic MR lymphography in prostate cancer patients

Oscar A. Debats, Midas Meijis, Geert J. S. Litjens, and Henkjan J. Huisman

Citation: *Medical Physics* **43**, 3132 (2016); doi: 10.1118/1.4951726

View online: <http://dx.doi.org/10.1118/1.4951726>

View Table of Contents: <http://scitation.aip.org/content/aapm/journal/medphys/43/6?ver=pdfcov>

Published by the American Association of Physicists in Medicine

Articles you may be interested in

Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method

Med. Phys. **41**, 092303 (2014); 10.1118/1.4893533

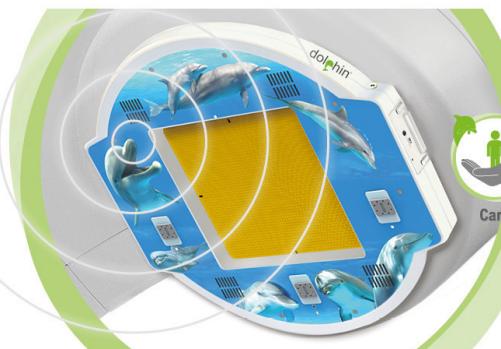
Inter-slice bidirectional registration-based segmentation of the prostate gland in MR and CT image sequences
Med. Phys. **40**, 123503 (2013); 10.1118/1.4829511

Automated 3-dimensional segmentation of pelvic lymph nodes in magnetic resonance images
Med. Phys. **38**, 6178 (2011); 10.1118/1.3654162

Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model
Med. Phys. **37**, 1579 (2010); 10.1118/1.3315367

Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information
Med. Phys. **35**, 1407 (2008); 10.1118/1.2842076

dolphin®
Patient QA
and Monitoring*
ONLINE READY!



iBa

*Useful for Pre-Treatment. Approval by Linac manufacturers for online use during patient treatment is pending.

Automated multistucture atlas-assisted detection of lymph nodes using pelvic MR lymphography in prostate cancer patients

Oscar A. Debats,^{a)} Midas Meijis, Geert J. S. Litjens, and Henkjan J. Huisman
Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen 6525 GA, The Netherlands

(Received 27 July 2015; revised 29 April 2016; accepted for publication 5 May 2016; published 31 May 2016)

Purpose: To investigate whether atlas-based anatomical information can improve a fully automated lymph node detection system for pelvic MR lymphography (MRL) images of patients with prostate cancer.

Methods: Their data set contained MRL images of 240 prostate cancer patients who had an MRL as part of their clinical work-up between January 2008 and April 2010, with ferumoxtran-10 as contrast agent. Each MRL consisted of at least a 3D T1-weighted sequence, a 3D T2*-weighted sequence, and a FLASH-3D sequence. The reference standard was created by two expert readers, reading in consensus, who annotated and interactively segmented the lymph nodes in all MRL studies. A total of 5089 lymph nodes were annotated. A fully automated computer-aided detection (CAD) system was developed to find lymph nodes in the MRL studies. The system incorporates voxel features based on image intensities, the Hessian matrix, and spatial position. After feature calculation, a GentleBoost-classifier in combination with local maxima detection was used to identify lymph node candidates. Multiatlas based anatomical information was added to the CAD system to assess whether this could improve performance. Using histogram analysis and free-receiver operating characteristic analysis, this was compared to a strategy where relative position features were used to encode anatomical information.

Results: Adding atlas-based anatomical information to the CAD system reduced false positive detections both visually and quantitatively. Median likelihood values of false positives decreased significantly in all annotated anatomical structures. The sensitivity increased from 53% to 70% at 10 false positives per lymph node.

Conclusions: Adding anatomical information through atlas registration significantly improves an automated lymph node detection system for MRL images. © 2016 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4951726>]

Key words: CAD, MR lymphography, prostate cancer, atlas registration, pelvic lymph node metastasis

1. INTRODUCTION

Approximately 220 000 men are diagnosed with prostate cancer (PCa) in the USA each year, and 27 000 die as a consequence of the disease. Those in whom the cancer has broken through the prostate capsule and formed metastases in pelvic lymph nodes have a substantially lower chance to survive.¹ Not only the prognosis but also the treatment options for a patient diagnosed with PCa depend on their lymph node status.

MR lymphography (MRL) remains the most accurate imaging modality for the assessment of metastatic involvement of pelvic lymph nodes.² MRL is MR imaging using a lymph node specific contrast agent based on ultrasmall superparamagnetic particles of iron oxide (USPIO), which result in signal intensity differences between metastatic and normal lymph node tissue.³ However, despite its high accuracy, MRL is currently not very widely used. Reasons for this are that the interpretation of MRL images is very time-consuming, with average reading times up to 80 min, and that readers with a very high level of experience are required for the assessment of the images.⁴

MRL interpretation time as well as dependence on high reader experience can be reduced by using a computer-aided detection (CAD) system. Such a system can detect the pelvic lymph nodes and present each one subsequently to the human reader, who can then assess more efficiently whether it is a metastatic lymph node or a healthy one, without having to search painstakingly through all regions of the pelvic area where lymph node metastases might be located. A modest number of studies have been published on the development of such a CAD system, most of them using CT images rather than MR images.^{5–7}

In a pilot study at our institution, it was shown that a generic CAD system based on pattern recognition can, after being properly trained, detect most lymph nodes from a set of pelvic MRL images. However, the challenge is to reduce the number of false positive (FP) detections, which currently is far too high for such a system to be usable in a routine clinical workflow.⁸ One of the issues is that false positives are not only generated in regions where lymph nodes can be expected to be found but also in areas of the MRI volume that are far away from all possible pathways of lymphogenic metastatic spread,

TABLE I. Scan parameters of the MRL sequences. TE: echo time; TR: repetition time.

| Description | Acronym | TE (ms) | TR (ms) | Flip angle (deg) | Matrix | Pixel size (mm) | Slice thickness (mm) |
|----------------------------|---------|------------|------------|---------------------|-----------|--------------------|-------------------------|
| T1-weighted spin echo | VIBE | 2.45 | 4.95 | 10 | 320 × 320 | 0.8 × 0.8 | 0.8 |
| T2*-weighted gradient echo | MEDIC | 11 | 20 | 10 | 320 × 320 | 0.8 × 0.8 | 0.8 |
| T2*-weighted | FLASH | 1.16 | 2.84 | 18 | 384 × 384 | 1.0 × 1.0 | 1.0 |

and even inside tissues that are not expected to contain any lymph nodes at all, such as muscle tissues, bone structures, or bowel loops.

Many false positive lymph node candidate detections can be dismissed by including anatomical information into the CAD system. When the system can make use of information defining which structures of the volume are actually pelvic bones, muscles such as the psoas and iliac muscle, intestinal structures, or major blood vessels, it cannot only exclude detections inside those structures but rather use their location to more accurately detect lymph nodes in the various lymph node regions of the pelvic area, which are defined by, and often even named after the surrounding anatomy.

In this study, we developed a CAD system prototype for detection of pelvic lymph nodes which includes anatomical information in the form of probabilistic atlas features. We compared the new CAD system to a generic pattern recognition based CAD prototype that did not include anatomical information, and evaluated the performance of both systems with respect to a database of lymph nodes annotated in MRL images.

2. METHODS

2.A. Summary of the CAD pipeline

In this section, the common CAD pipeline shared by both the generic CAD prototype and the atlas-enhanced CAD prototype is discussed. The prototypes are based on a pattern recognition system that is configured to work with MRL data sets containing T1-weighted (“VIBE”) and T2*-weighted (“MEDIC” and “FLASH”) image volumes

(Table I). Flowcharts of the CAD pipeline are shown in Figs. 1 and 2.

A set of image features are defined, which are implemented as feature filters, each of which uses one of the available MR images as input and has a feature map as output. After feature calculation, a voxel classification is performed which results in a lymph node likelihood between 0 and 1 for each voxel. For the voxel classification, a GentleBoost-classifier was used with regression stumps as weak learners. The number of weak learners was set to 200.

After voxel classification, a 3D likelihood map is obtained. Some example slices are shown in Fig. 3. On the likelihood map, we perform local maxima detection using a spherical window with a diameter of 10 mm, which is the maximum size of nonenlarged lymph nodes in MRL. After the local maxima detection, a merging step was performed to get rid of plateau-shaped maxima using connected component analysis. All components bigger than one voxel were reduced to only the voxel nearest to the center of gravity. The location and likelihood of the remaining local maxima define the lymph node detection points.

Once a list of points with high lymph node likelihood has been detected, it could be used directly in a clinical workflow, where the radiologist would be presented the list of detections in a dedicated viewing station and would be able to quickly go through the list, skipping over any nonlymph node detections, and appreciate true lymph nodes as normal or aberrant.

2.B. Features

Several types of voxel features were used in the voxel classification step of the CAD prototype.

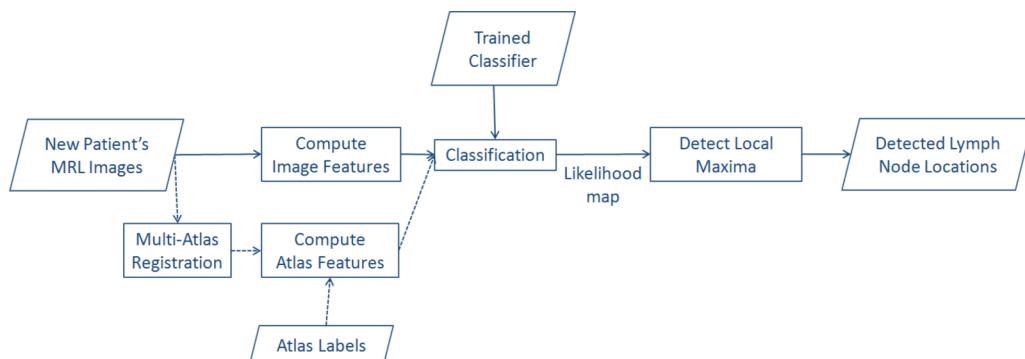


Fig. 1. Flowchart of the CAD pipeline. Solid arrows indicate the generic MRL CAD pipeline. Note that soft classification is performed, which does not output a hard class label but a likelihood value for each voxel. Dotted arrows indicate the added steps in the atlas-enhanced CAD pipeline.

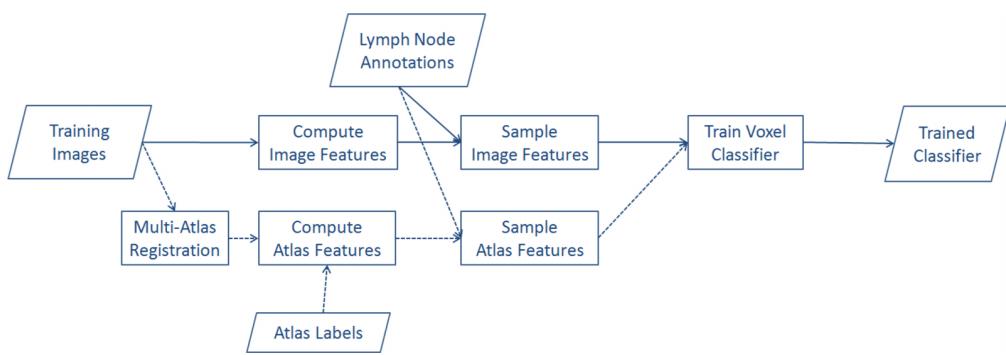


Fig. 2. Flowchart of the classifier training procedure. Solid arrows indicate the training procedure for the generic MRL CAD pipeline. After computing the feature maps, samples are taken from them at locations inside and outside the annotated lymph nodes and used as training data for the voxel classifier. Dotted arrows indicate the added steps in the atlas-enhanced CAD pipeline.

2.B.1. Image intensity

In MRL images, as with most MR sequences, the signal intensity I does not quantitatively reflect a physical property

of the tissue in the concerning voxel, and intensities can vary between scans. In a study by Wendt, it was shown that scaling the intensity values using first order statistics such as mean, median, and standard deviation can be a satisfactory and robust

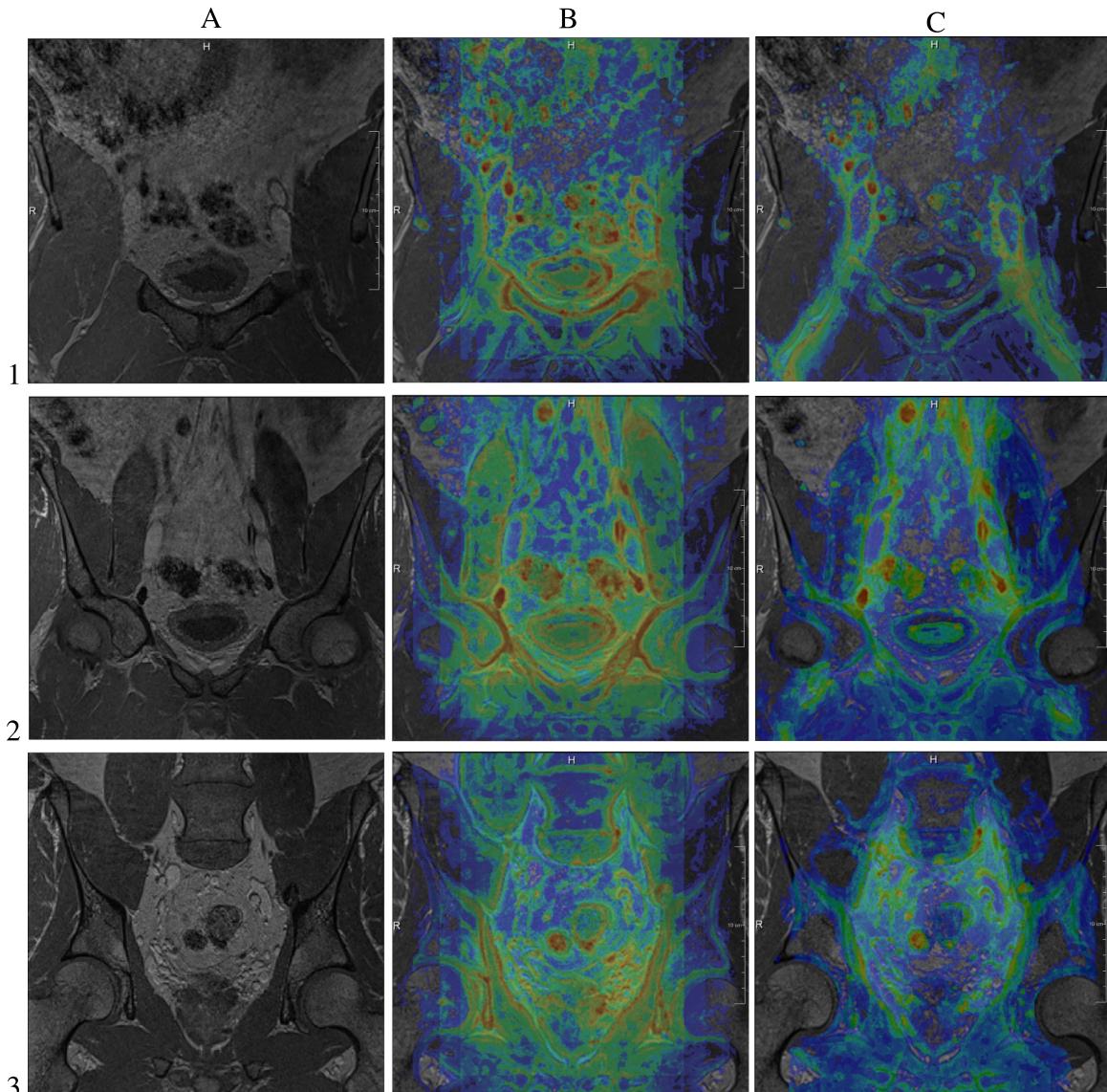


Fig. 3. Three example slices of the likelihood maps. (A) coronal view of the VIBE image volume; (B) likelihood map of the generic CAD prototype; (C) likelihood of the atlas-enhanced prototypes.

method of scaling image intensity.⁹ We applied scaling by the mean μ and standard deviation σ of the signal intensity within each image volume and included scaled intensity of the VIBE, MEDIC, and FLASH images as voxel features. $I_{\text{scaled}} = (I - \mu)/\sigma$.

2.B.2. Blobness, vesselness, and sheetness

One important characteristic of lymph nodes in MRL is that they appear as dark blob-shaped structures on the T1-weighted images. A common approach to detect blobs in medical images is by using a blob enhancement filter which determines the blob likelihood of a voxel \mathbf{x} using the eigenvalues $\lambda_{\sigma,k}$ of the Hessian matrix H_σ at scale σ , with $k = 1,2,3$ as described

$$P_{\text{blob}}(\mathbf{x}, \sigma) = \begin{cases} \frac{|\lambda_{\sigma,2}(\mathbf{x})| (|\lambda_{\sigma,2}(\mathbf{x})| - |\lambda_{\sigma,3}(\mathbf{x})|)}{|\lambda_{\sigma,1}(\mathbf{x})|} & \text{if } \lambda_{\sigma,k}(\mathbf{x}) < 0 \quad k \in 1,2 \\ 0 & \text{otherwise} \end{cases}. \quad (1)$$

Also sheet-like structures are included in the imaged volume, such as the cortical layers of pelvic bones, which often mimick the low signal intensities of lymph nodes in the T1-weighted images. Again, these should not generate a high lymph node likelihood themselves, so also the sheetness likelihood P_{sheet} was included in the feature set,

$$P_{\text{sheet}}(\mathbf{x}, \sigma) = \begin{cases} |\lambda_{\sigma,1}(\mathbf{x})| - |\lambda_{\sigma,2}(\mathbf{x})| & \text{if } \lambda_{\sigma,1}(\mathbf{x}) < 0 \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

2.B.3. Spatial position

Lymph nodes occur more often in certain areas, such as locations close to fatty tissue and around the main blood vessels. To enable the classifier to use this information, we included the craniocaudal, ventrodorsal, and left-right position relative to MR world coordinates. In our MRL studies, the VIBE and MEDIC images are acquired in a standardized procedure, such that in the craniocaudal direction they extend from the aortic bifurcation to the pubic bone, in the left-right direction they are centered on the median plane of the body, and in the ventrodorsal direction they extend from the abdominal wall to sacral vertebrae. Because of this standardized procedure, the origin of MR image space is linked to reference anatomy.

2.C. Probabilistic atlas features

We hypothesized that adding anatomical information to the CAD system can help in reducing the number of false positives. We chose multiatlas registration as strategy to encode the anatomical information as voxel likelihoods of belonging to a certain organ. Multiatlas registration is not

by Li et al.,¹⁰

$$P_{\text{blob}}(\mathbf{x}, \sigma) = \begin{cases} \frac{\lambda_{\sigma,3}(\mathbf{x})^2}{|\lambda_{\sigma,1}(\mathbf{x})|} & \text{if } \lambda_{\sigma,k}(\mathbf{x}) < 0 \quad k \in 1,2,3 \\ 0 & \text{otherwise} \end{cases}. \quad (1)$$

The Hessian eigenvalues can also be used to detect tube-like (or vessel-like) and sheet-like structures. Lymph nodes are connected to the lymph vessels which run alongside the blood vessels, so the lymph nodes are expected to be found in the neighborhood of the larger vessels, but the system should not give a high lymph node likelihood output for vessel-shaped structures themselves. To enable the classifier to exclude vessels, the feature set also includes the vesselness likelihood P_{vessel} ,

$$P_{\text{vessel}}(\mathbf{x}, \sigma) = \begin{cases} \frac{|\lambda_{\sigma,2}(\mathbf{x})| (|\lambda_{\sigma,2}(\mathbf{x})| - |\lambda_{\sigma,3}(\mathbf{x})|)}{|\lambda_{\sigma,1}(\mathbf{x})|} & \text{if } \lambda_{\sigma,k}(\mathbf{x}) < 0 \quad k \in 1,2 \\ 0 & \text{otherwise} \end{cases}. \quad (2)$$

widely used for abdominal MRI currently, but, for example, in brain MRI, it is common practice.¹¹

Ten MRL studies were randomly selected to serve as atlases, for each of which a reference segmentation was created as follows. The most important anatomical landmarks of the pelvic region were segmented interactively by one of the researchers (MM), using the medical image segmentation tool ITK-SNAP.¹² First, the osseous structures were segmented: femur, hip bones (ilium, pubis, and ischium), and spine, as illustrated in Fig. 4. Second, a number of intrapelvic and paraspinal structures were segmented: aorta, vena cava, iliac arteries and veins, psoas major muscles, iliocostalis muscles, and urinary bladder (Fig. 5). Also the gluteus maximus muscle was included. The remaining areas were labeled either as intra-abdominal space (which included the intrapelvic cavity)

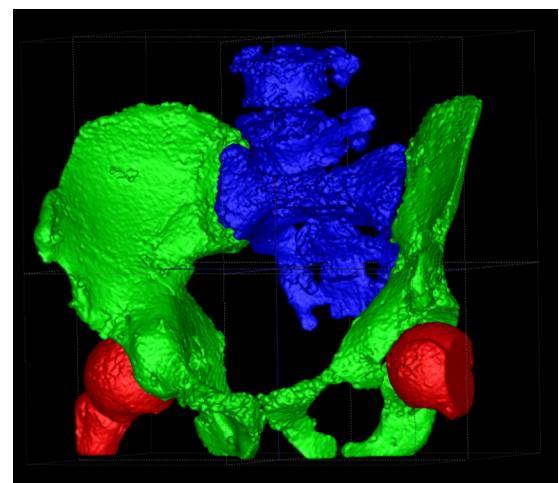


Fig. 4. A 3D view of the osseous structures in one of the atlases. Structures shown are the femur, pelvic bones, and spinal column.

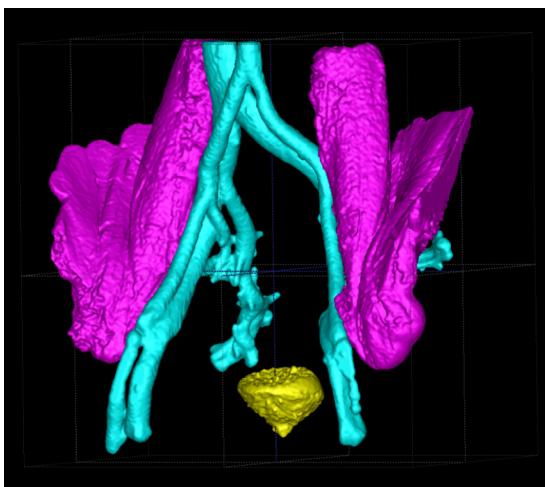


FIG. 5. A 3D view of one of the atlases, showing the main blood vessels, iliopsoas muscles, and urinary bladder.

or extrapelvic space. The VIBE images as well as the MEDIC and FLASH images were used for the interactive segmentation process.

For a new patient's MRL study to be processed by the CAD pipeline, the VIBE images of all ten atlases were deformably registered to the corresponding VIBE image of the new study, resulting in ten sets of transformation parameters, one for each atlas. To this end, we used the registration package ELASTIX, an open source software toolbox for rigid and nonrigid image registration.¹³ A two-step registration protocol was used. First, an affine rigid transformation was applied to account for coarse differences in the two images. After this, a nonrigid multiscale B-spline transformation was applied to align more subtle differences in pelvic anatomy between patients. Mutual Information was used as similarity measure, and optimization was performed using an adaptive stochastic gradient descent optimizer.¹⁴

Once the VIBE images of all atlases had been registered to the VIBE image of the new MR data set, the ten resulting sets of transformation parameters were used to transform the ten corresponding atlas label maps, to obtain ten segmentations for the new MR data set. Instead of using atlas label fusion to create binary segmentations, as can be done, for example, by majority voting, best atlas selection, or more advanced methods based on simultaneous truth and performance level estimation (STAPLE) or the selective and iterative method for performance level estimation (SIMPLE), the collection of atlases was treated as a nonparametric spatial probability distribution, and a frequency map was computed for each atlas label.¹⁵

The frequency map for each anatomical structure was defined as the 3D image in which each voxel scalar is the percentage of transformed atlas label maps in which the corresponding voxel is labeled with the concerning anatomical structure. The frequency maps were smoothed using a gaussian filter with a size of 1.95 mm (2.5 voxel lengths) to ensure spatial continuity and a smooth continuous scale. The smoothed frequency maps were used as features for the voxel classification.

2.D. Validation

2.D.1. Imaging and data collection

The two CAD prototypes (generic and atlas-enhanced) were evaluated using MRL images of a consecutive set of patients who fulfilled the following inclusion criteria:

- Biopsy-proven prostate cancer;
- ferumoxtran-10 MRL between January 2008 and April 2010;
- successful acquisition of VIBE, MEDIC, and FLASH image volumes.

A total of 240 patients were included in the analysis. Each patient underwent MR imaging enhanced with the USPIO-based lymph node specific contrast agent ferumoxtran-10 (Sinerem®, Guerbet, Paris, France) at the Radboud University Medical Center in Nijmegen, The Netherlands, as part of their clinical evaluation. Administration was done intravenously, 36–24 h before the MR imaging was performed. All patients received a drip infusion with a duration of approximately 30 min, containing a dose of 2.6 mg Fe per kg body weight. Immediately before the MR examination, Buscopan (20 mg i.v. and 20 mg i.m.) and Glucagon (20 mg i.m.) were administered in order to suppress bowel peristalsis. All patients provided informed consent for the use of their clinical images for research purposes. All imaging was performed using a 3.0 Tesla MR imaging system (Magnetom TrioTim; Siemens, Erlangen, Germany). Images were acquired in the coronal plane, covering the whole pelvis. Two MR series, called VIBE and MEDIC, respectively, were used in this analysis. The scan parameters are shown in Table I.

2.D.2. Reference standard

To assess the performance of the automated lymph node detection, reference lymph node annotations were created for the clinical evaluation of each patient's MRL, which was established in a consensus reading by two expert readers: an MD researcher specially trained in reading MRL scans (OD, 2 yr of MRL experience, >300 MRLs) and an abdominal radiologist (JB, >10 yr of MRL experience, >1000 MRLs).

The T1-weighted (VIBE) sequence, which is insensitive to USPIO contrast, was used for localization and assessment of shape and size of the LNs. Using this sequence, the individual lymph nodes were interactively segmented using the application Lymph Node Task Card, developed by Siemens, Malvern, PA, USA. The T2*-weighted (MEDIC) sequence was used to assess USPIO uptake for the clinical diagnosis of the patient, but was not used in our analysis.

2.D.3. Training data

The GentleBoost-classifier was trained with training sets containing feature data from randomly selected background voxels and lymph node voxels. Because lymph nodes are small organs, in their voxel representation after manual or interactive segmentation, their ratio of edge-voxels to nonedge voxels is

relatively high. Because of small segmentation errors, edge voxels but also voxels close to the edges are to be avoided when selecting training voxels. For reference lymph nodes with a long-axis diameter greater than 4 mm, we included in the training set only the voxels within a 4 mm radius of the center of gravity of the lymph node. Some lymph nodes in the reference segmentations had a volume less than 0.1 ml and were excluded from the training.

2.D.4. Registration accuracy

As the quality of the atlas features depends on accurate atlas registration, the accuracy of the registration (described in Sec. 2.C) was assessed. To do this, the MR images of the atlas studies were registered to the corresponding MR images of all other atlas studies. For each atlas study, the resulting nine transformation matrices were applied to the label maps of the other nine atlas studies. The nine resulting label maps were combined by majority voting. The combined label map could then be compared to the manually constructed one, in terms of spatial overlap. As overlap metric we used the Dice similarity coefficient (DSC),¹⁶ derived as a special case of the kappa statistic by Zijdenbos *et al.*,¹⁷

$$\text{DSC}_i = \frac{|\mathcal{S}_{R,i} \cap \mathcal{S}_{M,i}|}{\frac{1}{2}(|\mathcal{S}_{R,i}| + |\mathcal{S}_{M,i}|)}, \quad (4)$$

where $\mathcal{S}_{M,i}$ is the manually segmented voxel set for organ label i , and $\mathcal{S}_{R,i}$ is the voxel set for organ label i defined by the majority voting on the transformed label maps.

2.D.5. Visual inspection of the likelihood maps

The quality of the likelihood maps resulting from the voxel classification described in Sec. 2.A was evaluated visually, with special attention to areas of high likelihood in nonlymph node anatomical structures, such as muscles or bones, which will result in false positives if they have likelihoods in the same range as lymph nodes. A comparison was made between the likelihood maps resulting from the generic CAD prototype and those produced by the atlas-enhanced CAD prototype, to assess in a qualitative manner whether adding anatomical information to the system could be expected to yield a reduction of false positive detections.

2.D.6. False positive reduction

To quantify the reduction of false positives achieved by adding anatomical information, a histogram analysis of the likelihood of FP detection points, grouped by the anatomical structure in which they occurred, was performed. This was done by running the CAD system on the atlas studies, because for these studies, manually annotated label maps were available, defining the anatomical structures. A leave-one-case-out strategy was used to generate the atlas features for the atlas studies. Histograms with 100 bins were generated for the likelihood values of the FPs from both CAD prototypes in the labels “femur,” “pelvic bones,” “spinal column,”

“gluteus maximus,” “iliopsoas muscles,” “urinary bladder,” “blood vessels,” and “intestine.”

FPs with a likelihood value below the likelihood range of the true positive detection points do not affect the accuracy of a CAD system. In our case, likelihood values of 0.75 and below were irrelevant. Therefore, not only the median likelihood and the total number of FPs were compared between the two prototypes but also the number of FPs with likelihood values above 0.75. Since a normal distribution could not be expected, the statistical significance of the differences was tested using Wilcoxon’s rank sum test.¹⁸

2.D.7. Lymph node detection performance

To evaluate the performance of the two CAD prototypes, a free-response receiver operating characteristic (FROC) analysis was conducted, using a fivefold cross validation scheme. In each fold, 192 studies were used to train the classifier and 48 studies were used for evaluation. A detection point was considered a true positive if it was inside the boundaries of one of the reference lymph nodes; if not, it was considered a false positive. Each reference lymph node with a size of 0.1 ml or greater that was not “hit” by a detection point was considered a false negative.

2.D.8. Sensitivity analysis

As the number of atlases was a key design parameter of the atlas-enhanced CAD prototype, a sensitivity analysis was performed to assess its relative contribution to the performance of the system. For this purpose, a random subset of 20 MRL studies was selected, and the system was trained and evaluated as described above, but with different sets of atlas features, constructed using either 1, 3, 5, or 10 atlases.

3. RESULTS

3.A. Registration accuracy

The results of the analysis of registration accuracy are summarized in Table II. Registration proved to be more accurate for some anatomical structures than for others. Of the structures analyzed, the femur had the highest accuracy (mean DSC 0.868), followed by the other musculoskeletal structures. For the blood vessels, a lower mean DSC was achieved, and the intestinal structures show the lowest value (0.155). The standard deviation over the ten randomly selected atlas studies was lowest for the musculoskeletal structures and the blood vessels (range 0.03–0.06). The highest standard deviation was found for the urinary bladder (0.17). An example of a registered pelvic anatomy atlas is presented in Fig. 6, which shows that a good visual alignment was achieved for the musculoskeletal structures.

3.B. Visual inspection of the likelihood maps

Three example slices of VIBE images with their likelihood maps as overlay are shown in Fig. 3. As can be seen, the

TABLE II. Registration accuracy. The highest Dice similarity is achieved for the osseous structures in general and, more specifically, the femur. μ : mean; σ : standard deviation.

| | DSC | |
|-------------------|-------|----------|
| | μ | σ |
| Femur | 0.868 | 0.05 |
| Pelvic bones | 0.773 | 0.04 |
| Spinal column | 0.741 | 0.06 |
| Gluteus maximus | 0.816 | 0.03 |
| Iliopsoas muscles | 0.831 | 0.03 |
| Urinary bladder | 0.781 | 0.17 |
| Blood vessels | 0.544 | 0.05 |
| Intestine | 0.155 | 0.10 |

likelihood map of the generic CAD prototype [Fig. 3(B)] contains a number of high likelihood areas in anatomical structures such as the cortex of the pubic bones and the surrounding muscles. In the likelihood map of the atlas-enhanced CAD prototype [Fig. 3(C)], these areas are much less abundant.

3.C. False positive reduction

The histograms showing the likelihood distribution of the FPs of both CAD prototypes are displayed in Fig. 7. The results of the histogram analysis are summarized in Tables III and IV. The atlas-enhanced prototype achieves a strong reduction of FPs compared to the generic prototype, as shown by the reduction of detection points inside almost all anatomical structures in the reference segmentations. The urinary bladder, which shows an increase of the number of

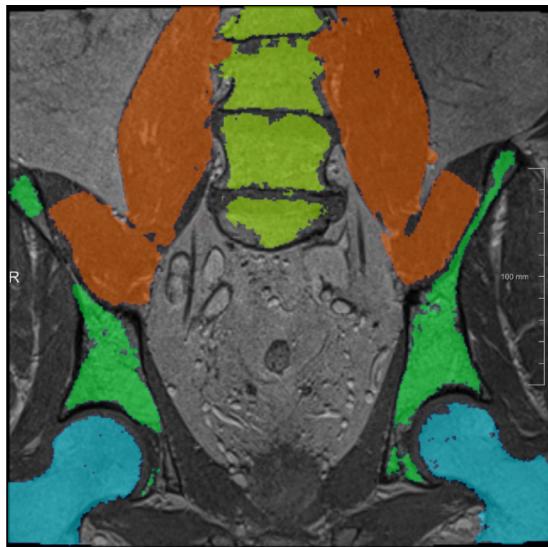


FIG. 6. Example of a registration result, showing that a good visual alignment was achieved. The gray scale image is a coronal example slice of the VIBE image of one of the MRL data sets. The color overlay shows a label map generated by registration of one of the atlases to the VIBE image. Anatomical structures shown are the femur (blue), pelvic bones (green), spinal column (yellow), and iliopsoas muscles (orange).

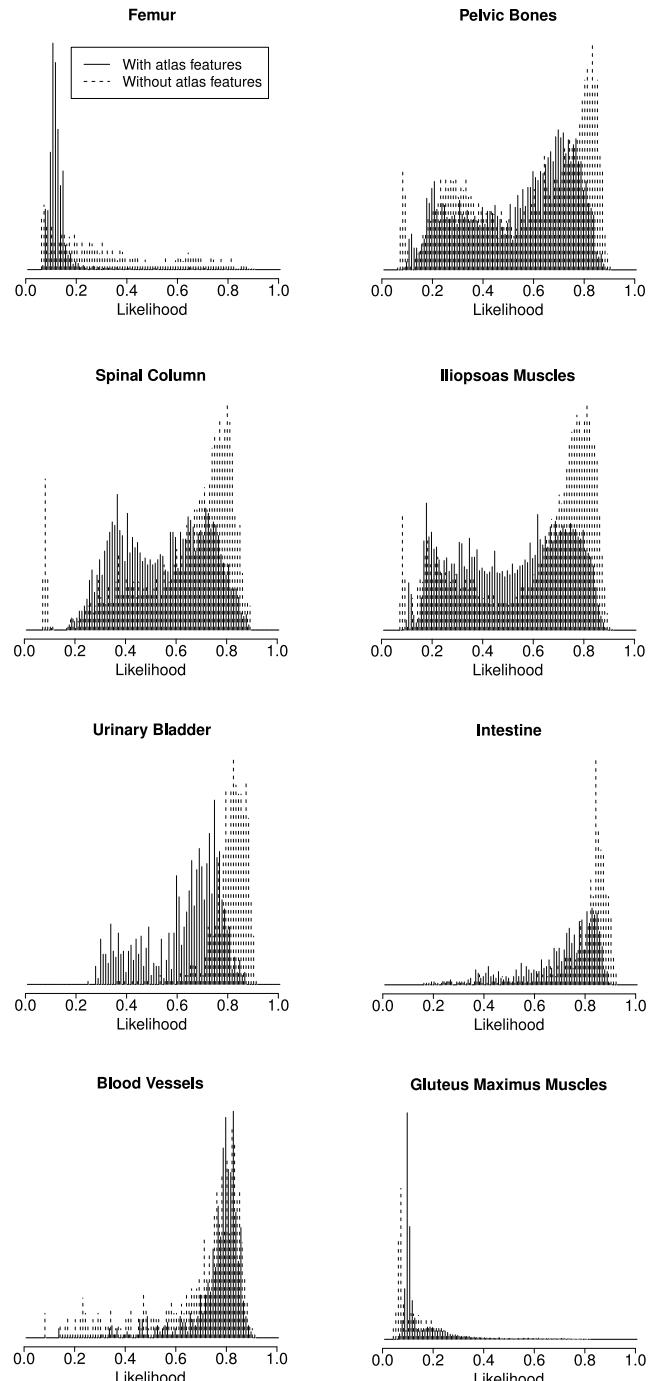


FIG. 7. Histograms of false positives in various anatomical structures. For most anatomical structures (femur, pelvic bones, spinal column, iliopsoas muscles, urinary bladder, and intestine), the addition of atlas-based anatomical information strongly reduces the number of false positives in the high likelihood (>0.8) domain. In the gluteus maximus muscles, almost no false positives with high likelihoods are generated. The only type of anatomical structure where high likelihood false positives are not substantially reduced, are the blood vessels.

FPs, is the only exception. The median likelihood value of the false detection points was significantly decreased in all anatomical structures except for the gluteus maximus and the blood vessels. (Table III). The number of FPs with likelihood values above a threshold of 0.75 is strongly reduced in most

TABLE III. Reduction of FPs. The number of FPs is strongly reduced in all anatomical structures except for the urinary bladder. The median likelihood value of the FPs is significantly decreased in all anatomical structures except for the gluteus maximus and the blood vessels.

| | No. of FPs/scan | | Median likelihood | | <i>p</i> |
|-------------------|-----------------|----------------|-------------------|----------------|----------|
| | Generic | Atlas-enhanced | Generic | Atlas-enhanced | |
| Femur | 187 | 160 | 0.256 | 0.114 | <0.001 |
| Pelvic bones | 649 | 539 | 0.630 | 0.592 | <0.001 |
| Spinal column | 320 | 295 | 0.692 | 0.554 | <0.001 |
| Gluteus maximus | 950 | 919 | 0.090 | 0.101 | <0.001 |
| Iliopsoas muscles | 940 | 824 | 0.693 | 0.533 | <0.001 |
| Urinary bladder | 92 | 109 | 0.814 | 0.669 | <0.001 |
| Blood vessels | 111 | 75 | 0.753 | 0.791 | <0.001 |
| Intestine | 107 | 87 | 0.832 | 0.747 | <0.001 |

anatomical structures. The number of thresholded FPs in the gluteus maximus was paradoxically increased, although their absolute number, at 2.3 per scan, remained fairly low. The number of thresholded FPs in the blood vessels remained essentially the same. In all other anatomical structures, the number of thresholded FPs was reduced by at least 48.8% (Table IV).

3.D. Detection performance

The results of the FROC analysis are summarized in Figs. 8 and 9. The atlas-enhanced prototype shows a strong improvement of detection performance. The maximum sensitivity is similar for both prototypes, approximately 96%. However, in the sensitivity range from 20% to 80%, the number of false positives is reduced by a factor 2.5. For example, at 80% sensitivity, the nonenhanced CAD prototype produces 43.4 FPs/lymph node (509 FP/scan). With the atlas-enhanced prototype, the number is reduced to 17.6 FP/LN (207 FP/scan). The sensitivities at a FP level of 10 FP/LN (117 FP/scan) were 0.703 and 0.531 for the systems with and without atlas-based anatomical information, respectively.

TABLE IV. Reduction of FPs with likelihood values above 0.75. The number of FPs is strongly reduced in all anatomical structures except for the gluteus maximus and the blood vessels. The number of FPs in the gluteus maximus was paradoxically increased, although their absolute number, at 2.3 per scan, remained fairly low. The number of FPs in the blood vessels remained essentially the same.

| | No. of FPs/scan | | |
|-------------------|-----------------|----------------|--------------|
| | Generic | Atlas-enhanced | % difference |
| Femur | 9.9 | 0.2 | -98.0 |
| Pelvic bones | 204 | 88.3 | -56.7 |
| Spinal column | 108 | 39 | -63.9 |
| Gluteus maximus | 0.7 | 2.3 | +229 |
| Iliopsoas muscles | 335 | 129 | -61.5 |
| Urinary bladder | 77.2 | 20.9 | -72.9 |
| Blood vessels | 56.5 | 54.3 | -3.9 |
| Intestine | 84.4 | 43.2 | -48.8 |

3.E. Sensitivity analysis

The results of the sensitivity analysis are summarized in the FROC plots in Fig. 10. The plots reveal that the detection performance improves with the number of atlases but levels off at 5 atlases. Adding more atlases does not improve the performance [Fig. 10(A)].

Varying the choice of atlases did not have a substantial influence on CAD performance. An example of this is shown in Fig. 10(B), where the FROC curves are plotted for three systems, each trained with atlas features constructed using a different set of three randomly selected atlases.

4. DISCUSSION

Our results showed that adding anatomical information in the form of probabilistic multitalas features improved the CAD prototype by reducing the number of FPs. The generic and atlas-enhanced CAD prototypes were evaluated on a large number of lymph nodes, for all of which a reference

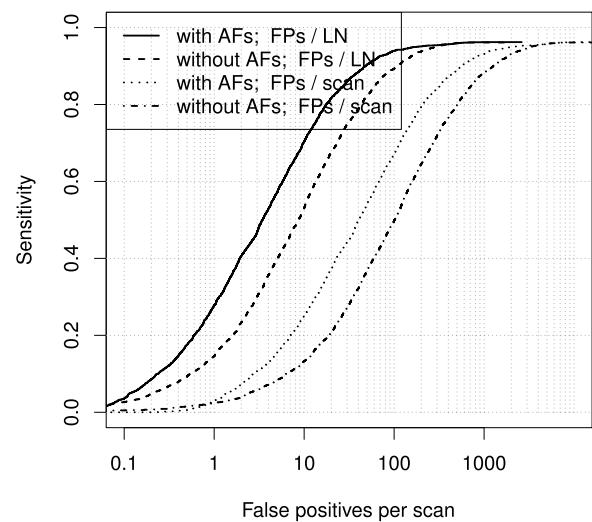


FIG. 8. FROC plot of detection performance. AFs: atlas features; FPs: false positives; LN: lymph node. At 80% sensitivity, the nonenhanced CAD prototype produces approximately 40 FPs/lymph node. With the atlas-enhanced prototype, the number of FPs is reduced by a factor 2.5.

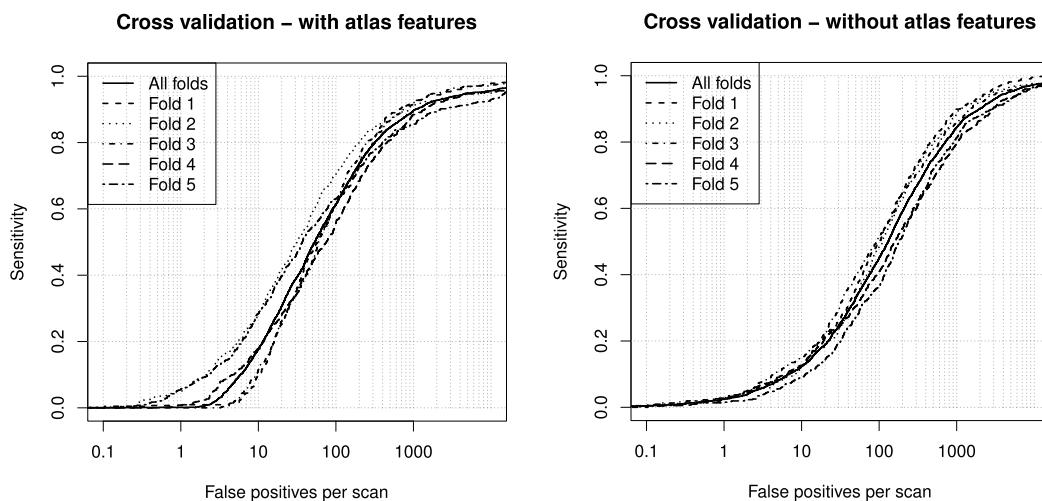


FIG. 9. Variability analysis of detection performance. To visualize the variability, separate FROC curves are plotted for each of the folds of the fivefold cross validation.

segmentation was available, created in consensus by two expert readers who annotated all MRL data sets that met the inclusion criteria in a consecutive set of patients. A total of 240 MRLs was included, all of which were successfully processed by the two CAD prototypes. The registration procedure proved to be robust as it achieved an acceptable result for all 240 MRLs. A result was considered to be acceptable when it contained no serious misregistrations on visual inspection, as was the case for all MRL studies except for the fact that the exact shape of the intestines in the target study could generally not be captured by the registration procedure. However, this is considered to be inevitable, due to the large differences in shape and location of the intestinal structures between individuals.

A satisfactory registration accuracy was achieved for most anatomical structures. This was appreciated by visual inspection of the registered atlases and confirmed by the Dice overlap values for almost all pelvic anatomical structures that

were labeled in the atlases. The accuracy was higher for some organs than for others, as shown in Table II. Partly, this can be explained by the fact that some structures (e.g., intestine) are more challenging in registration than other structures (e.g., bone, muscle), because they show more variation in location between subjects. The size and shape of an organ may not only influence the registration process itself but will also determine how much the DSC value decreases for certain types of misregistration. For example, blood vessels, being elongated structures, will lose relatively much Dice overlap when the registration result is a little off in the transverse plane.

We assumed that in general, for any one of the anatomical structures studied, a higher registration accuracy would lead to a better false positive reduction. However, this was not a topic of investigation in our study. Note that we did not assume a correlation between registration accuracy and FP reduction across different anatomical structures.

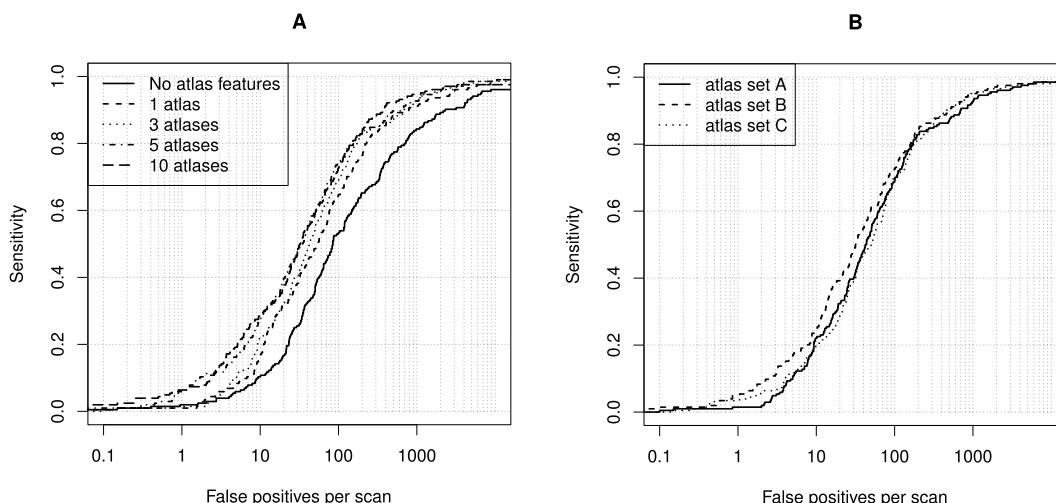


FIG. 10. (A) FROC plot of the sensitivity analysis of the number of atlases used to build the probabilistic atlas features. The differences in detection performance while varying the number of atlases are relatively small. Although the curve for the 10-atlas system lies clearly above the curve for the 1-atlas system, the performance with 5 and with 10 atlases is essentially the same. (B) FROC plot of CAD systems with atlas features based on three atlases: three sets of three randomly selected atlases were used to build the probabilistic atlas features and train a CAD system.

While FP reduction and detection performance were evaluated on all 240 MRL studies, registration accuracy was evaluated on the ten atlas studies, because these were the studies for which manual segmentations of the anatomical structures were available. The ten atlas studies were randomly selected, and the variation in registration accuracy between them was low (see the σ values in Table II), indicating that they are sufficiently representative as a subset.

Our registration results are comparable to the scarce literature on pelvic atlas registration. A number of studies have been published on automatic prostate segmentation and registration,^{19–21} but since many of the subjects in our population have undergone radical prostatectomy, the prostate is not one of the organs included in our atlases. Only a few studies are available in which similar organs are segmented. In the work of Acosta *et al.*,²² a DSC value of 0.58 is reported for the urinary bladder. Dowling *et al.*²³ report a slightly higher value of 0.64, which is lower than our result (DSC 0.781). Their Dice overlap for pelvic bones (DSC 0.79) is comparable to our value (0.773). Also Will *et al.*²⁴ report a Dice overlap for the pelvic bones (0.719) that is similar to our result.

The likelihood maps were strongly improved by the addition of anatomical information. Visual inspection of the likelihood maps indicated that, particularly inside muscle tissue and bone cortex, many high-likelihood areas that were present in the output of the generic prototype were not visible in the likelihood maps of the atlas-enhanced system. In a histogram analysis of the likelihood values of the detected lymph node candidates, it was shown that adding the atlas-based anatomical information resulted in a significant decrease of the median likelihood of the false positive detections in most anatomical structures. False positives with likelihood values above 0.75 were reduced by a factor two or more in six of the eight annotated structures in the atlases.

The atlas-enhanced CAD prototype reduced the number of false positives by a factor 2.5. At a sensitivity level of 80%, the generic system showed 43.4 false positives per lymph node, while the enhanced system reached the same sensitivity at only 17.6 FPs/LN. Both systems were able to find 96% of the reference lymph nodes. Closer inspection of the 1% highest-likelihood false positive detections revealed that some of them were actually true lymph nodes that were missing in the reference annotations. This means that our results may have underestimated the true performance of the CAD prototypes.

The sensitivity analysis showed that a too low number of atlases (1 or 3) used for the construction of the probabilistic atlas features negatively affected CAD performance. We conclude that at least five atlases are required.

A limited number of studies have been published on automated lymph node detection, and almost all of them have used computed tomography (CT) data rather than MR data. In 2007, Kitasaka *et al.* of Nagoya University, Japan, presented a system for lymph node detection in CT images by extracting blob-like structures using a 3-D minimum directional difference filter.⁵ They applied their method to only five cases of 3D abdominal CT images. For lymph nodes with a diameter of 5 mm or more, they report a sensitivity of

0.89 at a false positive (FP) rate of 17 000 FP's/case, before FP reduction. After a number of different FP reduction steps, the performance was increased to a FP rate of 58 FP's/case at a remaining sensitivity of 0.57. In 2009, the same group presented a method for detection of mediastinal lymph nodes, based on Hessian eigenvalue analysis. Evaluating on a set of five contrast-enhanced chest CT datasets, they report an average sensitivity of 82.1%, with a PPV of 13.3% and a FP rate of 113.4 FP/case.

To the best of our knowledge, the study by Yan *et al.* is the only published study on automated lymph node detection in MR images.²⁵ They combined mean shift clustering with region-based pattern recognition using a support vector machine (SVM) to detect the lymph nodes. They evaluated their method on 14 MRL cases, and report an average positive predictive value (PPV) of 0.64, which in their text is somewhat confusingly called "specificity." However, this value does not take into account the lymph nodes in their "unsure" category which, when included, reduces the PPV to 0.60. The FP rates, excluding and including the unsure detections, are 6.5 FP/scan and 8.1 FP/scan respectively. Unfortunately, no sensitivity was reported and no info on the number of missed lymph nodes was given, which makes it difficult to interpret their results. After all, the merit of any FP rate (or specificity) can only be appreciated in relation to the corresponding sensitivity.

At a sensitivity of 80%, our atlas-enhanced prototype might already be useful when combined with automated segmentation of the lymph node candidates, as described in previous work,²⁶ helping reduce search time for the human observer. However, a lower number of FPs would be preferred for clinical application.

This study has some limitations. The atlases contained a relatively low number of separate labels. Adding more anatomical structures to obtain a more detailed atlas would provide the classifier with more anatomical information, which might improve classification performance. Atlas registration worked poorly for some structures: for blood vessels as well as intestines, the DSC values were lower than expected. For the intestines, this might be due to the fact that these structures have a high variation in position and shape among patients. The position and shape of the segmented blood vessels do not vary so much, but may still be the cause of their poor Dice overlap: as they are elongated and relatively thin compared to voxel size, a registered atlas that is slightly shifted in the plane perpendicular to the vessel, will induce a large decrease in DSC value. Another limitation is that the CAD prototypes included only voxel-based classification. A more elaborate CAD pipeline, including also a region-based classification step, would be able to further improve performance.

In the future research, we aim to follow two strategies: improving the anatomical information, and extending the CAD prototype by adding region segmentation and classification. Improvement of the anatomical information, by improving the anatomy features, may further reduce the number of false positive lymph node detections and may also increase sensitivity. This might be achieved by applying more advanced methods of atlas label fusion, and/or by

combining atlas techniques with other segmentation methods. Such an approach would be valuable especially for anatomical structures for which the atlas method is not optimal, such as the blood vessels and the intestines. Extending the CAD prototype by adding a region classification stage can further reduce FPs by improving discrimination of true lymph nodes from other blob-like structures.

In conclusion, we developed a prototype CAD system that detects 96% of the lymph nodes visible in MRL images. We showed that adding anatomical information can reduce FPs significantly and thereby improve performance.

ACKNOWLEDGMENT

This project was funded by Grant No. KUN2007-3971 from the Dutch Cancer Society. The content is solely the responsibility of the authors and does not necessarily represent the official views of this institution.

CONFLICT OF INTEREST DISCLOSURE

The authors have no relevant conflicts of interest to disclose.

^aElectronic mail: debats@gmail.com

¹R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *Ca-Cancer J. Clin.* **65**(1), 5–29 (2015).

²A. S. Fortuin, R. J. Smeenk, H. J. M. Meijer, A. J. Witjes, and J. O. Barentsz, "Lymphotropic nanoparticle-enhanced MRI in prostate cancer: Value and therapeutic potential," *Curr. Urol. Rep.* **15**, 389 (7pp.) (2014).

³R. A. M. Heesakkers, A. M. Hövels, G. J. Jager, H. C. M. van den Bosch, J. A. Witjes, H. P. J. Raat, J. L. Severens, E. M. M. Adang, C. Hulsbergen van der Kaa, J. J. Fütterer, and J. Barentsz, "MRI with a lymph-node-specific contrast agent as an alternative to CT scan and lymph-node dissection in patients with prostate cancer: A prospective multicohort study," *Lancet Oncol.* **9**, 850–856 (2008).

⁴H. C. Thoeny, M. Triantafyllou, F. D. Birkhaeuser, J. M. Froehlich, D. W. Tshering, T. Binser, A. Fleischmann, P. Vermathen, and U. E. Studer, "Combined ultrasmall superparamagnetic particles of iron oxide-enhanced and diffusion-weighted magnetic resonance imaging reliably detect pelvic lymph node metastases in normal-sized nodes of bladder and prostate cancer patients," *Eur. Urol.* **55**, 761–769 (2009).

⁵T. Kitasaka, Y. Tsujimura, Y. Nakamura, K. Mori, Y. Suenaga, M. Ito, and S. Nawano, "Automated extraction of lymph nodes from 3-D abdominal CT images using 3-D minimum directional difference filter," in *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science Vol. 4792 (Springer, New York, NY, 2007), pp. 336–343.

⁶H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," in *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science Vol. 8673 (Springer, New York, NY, 2014), pp. 520–527.

⁷A. Seff, L. Lu, K. M. Cherry, H. R. Roth, J. Liu, S. Wang, J. Hoffman, E. B. Turkbey, and R. M. Summers, "2D view aggregation for lymph node detection using a shallow hierarchy of linear classifiers," in *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science (Springer, New York, NY, 2014), Vol. 8673, pp. 544–552.

⁸M. Meijs, O. A. Debats, and H. J. Huisman, "The evaluation of multi-structure, multi-atlas pelvic anatomy features in a prostate MR lymphography CAD system," *Proc. SPIE* **9414**, 94140T (2015).

⁹R. Wendt 3rd, "Automatic adjustment of contrast and brightness of magnetic resonance images," *J. Digital Imaging* **7**, 95–97 (1994).

¹⁰Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med. Phys.* **30**, 2040–2051 (2003).

¹¹M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Methods Programs Biomed.* **104**, e158–e177 (2011).

¹²P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage* **31**, 1116–1128 (2006).

¹³S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "ELASTIX: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**, 196–205 (2010).

¹⁴S. Klein, J. P. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *Int. J. Comput. Vision* **81**(3), 227–239 (2009).

¹⁵X. Artaechevarria, A. Muñoz-Barrutia, and C. O. de Solórzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imaging* **28**, 1266–1277 (2009).

¹⁶L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**, 297–302 (1945).

¹⁷A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Trans. Med. Imaging* **13**, 716–724 (1994).

¹⁸F. Wilcoxon, S. Katti, and R. A. Wilcox, "Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test," in *Selected Tables in Mathematical Statistics* (American Mathematical Society, Providence, RI, 1970), Vol. 1, pp. 171–259.

¹⁹S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Med. Phys.* **35**, 1407–1417 (2008).

²⁰G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-aided detection of prostate cancer in MRI," *IEEE Trans. Med. Imaging* **33**, 1083–1092 (2014).

²¹W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Dual optimization based prostate zonal segmentation in 3D MR images," *Med. Image Anal.* **18**, 660–673 (2014).

²²O. Acosta, A. Simon, F. Monge, F. Commandeur, C. Bassirou, G. Cazoulat, R. de Crevoisier, and P. Haigron, "Evaluation of multi-atlas-based segmentation of CT scans in prostate cancer radiotherapy," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, New York, NY, 2011), pp. 1966–1969.

²³J. A. Dowling, J. Lambert, J. Parker, O. Salvado, J. Fripp, A. Capp, C. Wratten, J. W. Denham, and P. B. Greer, "An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **83**, e5–e11 (2012).

²⁴S. Will, P. Martirosian, C. Würslin, and F. Schick, "Automated segmentation and volumetric analysis of renal cortex, medulla, and pelvis based on non-contrast-enhanced T1- and T2-weighted MR images," *MAGMA* **27**, 445–454 (2014).

²⁵M. Yan, Y. Lu, R. Lu, M. Requardt, T. Moeller, S. Takahashi, and J. Barentsz, "Automatic detection of pelvic lymph nodes using multiple MR sequences," *Proc. SPIE* **6514**, 65140W (2007).

²⁶O. A. Debats, G. J. S. Litjens, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, "Automated 3-dimensional segmentation of pelvic lymph nodes in magnetic resonance images," *Med. Phys.* **38**, 6178–6187 (2011).