



Automated vertebrae localization and identification by decision forests and image-based refinement on real-world CT data

Ana Jimenez-Pastor¹ · Angel Alberich-Bayarri^{1,2} · Belen Fos-Guarinos¹ · Fabio Garcia-Castro¹ · David Garcia-Juan¹ · Ben Glocker³ · Luis Marti-Bonmati^{1,4}

Received: 30 May 2019 / Accepted: 5 September 2019 / Published online: 14 September 2019
© Italian Society of Medical Radiology 2019

Abstract

Purpose Development of a fully automatic algorithm for the automatic localization and identification of vertebral bodies in computed tomography (CT).

Materials and methods This algorithm was developed using a dataset based on real-world data of 232 thoraco-abdominopelvic CT scans retrospectively collected. In order to achieve an accurate solution, a two-stage automated method was developed: decision forests for a rough prediction of vertebral bodies position, and morphological image processing techniques to refine the previous detection by locating the position of the spinal canal.

Results The mean distance error between the predicted vertebrae centroid position and truth was 13.7 mm. The identification rate was 79.6% on the thoracic region and of 74.8% on the lumbar segment.

Conclusion The algorithm provides a new method to detect and identify vertebral bodies from arbitrary field-of-view body CT scans.

Keywords Artificial intelligence · Decision forest · Spine · Computed tomography

Introduction

In clinical routine practice dealing with spinal abnormalities and pathologies, the localization and identification of the vertebral bodies is a crucial step for an appropriate clinical diagnosis, surgical planning and follow-up assessment. This task is time-consuming, hindering the radiologists' workflow. Manual labeling and measuring of all vertebrae are frequently performed to calculate the vertebra height ratios in order to evaluate fractures and to determine the CT-derived bone mineral density when dealing with osteoporosis.

The first approaches developed with these purposes were thought from a semi-automatic view [1–3]. Despite they did not require a high computational burden, they required some anatomical landmarks.

The main difficulties to create a fully automated system to robustly locate and identify the vertebral bodies in CT images are related to the similarity among anatomical landmarks of the spinal column, variability in spine curvature and shape, possible artifacts caused by metal implants, the presence of different bone abnormalities and pathologies and restrictions in the *z*-axis field of view (FOV) with acquisitions not covering all vertebrae in the longitudinal axis, based on the specific region of interest being studied. Additionally, the use of real-world data (RWD), that is, the use of studies coming directly from the hospital image repository, acquired with the inherent variability and biases of daily practice conditions, is a key point to develop a methodology directly applicable in clinical routine.

Some previous methods were focused on specific regions of the spine [4–6] or relied on prior knowledge about which region was examined and visible [7]. Therefore, these might not be able to be used for general universal applications where no assumptions of the visible region are made.

✉ Ana Jimenez-Pastor
anajimenez@quibim.com

¹ QUIBIM SL, Valencia, Spain

² Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain

³ Computing Department, Imperial College London, London, UK

⁴ Radiology Department, La Fe Polytechnics and University Hospital, Valencia, Spain

There are some other methods that are applied to arbitrary FOVs; however, these methods are based on mathematical models for vertebrae localization [8, 9]. The main problem with these methods relies on those abnormal cases, which make difficult the labor of implementing a model on which all the variability in both vertebra's shape and appearance among population is taken into account.

In recent years, some learning-based methods, applied to arbitrary FOV CT images, have been developed. Glocker et al. [10] proposed a supervised machine learning method based on random regression forests (RRF) combined with a refinement step based on hidden Markov models. However, some problems related with the narrow FOV on those pathological cases were found. For that reason, Glocker et al. [11] proposed a new method based on random-classification forests, which allowed obtaining higher performance on abnormal and pathological cases. In 2015, Suzani et al. introduced a similar methodology by using the same image features extraction steps as in [10, 11]; however, a novelty on the classification task was introduced, and it was used a feed-forward deep neural network (DNN). However, DNN does not take advantage of all the spatial information contained within the images as the convolutional neural networks (CNN) do. A DNN needs a prior feature extraction step; however, CNN is designed to include this feature extraction on its architecture. Prior to the fully connected layers, with classification purposes, there are several layers based on convolutional filters extracting from very simple features, such as brightness and edges, to most complex features that uniquely define the image. Some algorithms based on CNN have also been proposed for the automatic localization and identification of vertebrae in spine CT. Chen et al. [13] introduced a hybrid method based on the combination of a random forest classifier to roughly locate vertebra candidates with a joint convolutional neural network (J-CNN) for a more accurate vertebra localization. Yang et al. [14] developed a method based on a deep image-to-image network (DI2IN) to initialize vertebra locations combined with a sparsity regularization refinement step. Recently, Liao et al. [15] proposed a method which combined a 3D fully convolutional neural network (FCN) to extract short-range contextual information around the target vertebra, with a bidirectional recurrent neural network (Bi-RNN) to extract long-range contextual information to encode the spatial and contextual information among the vertebrae of the whole FOV.

Machine learning (ML) applications, as a branch of artificial intelligence (AI), have grown significantly in the last decade. Decision forests are a supervised ML technique composed by decision trees, the word supervised means that an associated set of output data is needed for each set of training data. Decision trees are known to suffer from over-fitting (high fitting to training data but poor predictive

performance). In order to minimize high fitting bias, the parameters of each split node of the forest are optimized only over a randomly sampled subset of all possible features. Not only that, the random sampling together with the ensemble of many trained decision trees yields a much better generalization. A significant advantage of ML over other AI techniques, such as deep learning (DL), is that they do not require high computational loads when training a model, and additionally, they offer better performance when leading with a small training dataset.

We aimed to propose a novel method for vertebrae centroid localization and identification on CT images. The developed method will be based on a two-stage approach, combining supervised learning by random decision forests [16] with image processing techniques. The method might be able to predict the vertebral bodies position present on CT exams where no assumptions about the scanned region is made.

Materials and methods

Dataset

The dataset was collected retrospectively through an observational study approved by the Ethics Committee and waived from informed consent collection. The data finally included and used for the development of the algorithm consisted of 232 multi-detector CT scans acquired both with 64 and 256 detector systems (Philips CT Brilliance and iCT, Best, The Netherlands) with different arbitrary field-of-views. The population series were patients that underwent either thoracic, abdominopelvic or cervical-thoracic-abdominopelvic CT examinations in a single longitudinal continuous acquisition in a period of 12 months (May-2015 to May-2016), including patients between 18 and 80 years old. In order to reach the goal of the study, no pathological conditions were excluded to enrich the algorithm development process, and patients with spinal pathologies such as scoliosis or vertebral fusion were included.

All the reconstructed images had a matrix size of either 512×512 or 768×768 with a pixel spacing ranging from 0.55 to 0.97 mm². The number of slices in each volume varied from 184 to 1629, with a slice thickness ranging from 0.5 to 3 mm.

This dataset was split into two separate groups, and cases were randomly distributed, using 80% (186 CT scans) to train and 20% (46 CT scans) to test.

Centroid annotation

All CT volumes were reconstructed in the coronal and sagittal orientations for annotation. The labeling was done, by

an expert in the radiological field, by selecting the centroids of all vertebrae present on the images. The set of annotated vertebrae was defined as $C = \{T1, \dots, T12, L1, \dots, L5, S1\}$, which contained both whole thoracic and lumbar regions and one additional sacrum vertebra.

For each image, the annotated centroids were stored in a matrix which included the absolute coordinates ($c_i \in \mathbb{R}^3$) and the specific label of each vertebra (C_i) present on the image. All these images were manually annotated by a radiology expert using an application designed ad hoc.

Methodology

The method was developed using both Python 3.5 and MATLAB r2016a (Mathworks Inc., Natick MA, USA) in a scientific computing server with an Intel i7 processor running at 3.6 GHz and 54 Gb of RAM memory.

The approach was developed on two stages by combining RRF with image-based algorithms. The first stage aims to detect all vertebrae centroid positions within the CT exam using a learning-based decision forests method. The second phase points to refine the prior detection considering the spine morphology by obtaining the spinal cord position using voxel-wise operations.

Detection based on random regression forests

An initial approach to locate the centroids position of all the vertebral bodies present in the images was performed by training a RRF network. It was trained a single RRF for all vertebral centroids present on an image.

For the image labeling phase, an in-house software application was developed to visualize and label the centroids of all vertebrae in the datasets of the study. Regarding the vertebrae localization and identification problem, intensity-based features were used as training input data ($f_i \in \mathbb{R}^d$). F features were extracted from each randomly selected voxel, and the distances from each randomly selected voxel to each annotated vertebra centroid were used as training output data. The goal was to learn a mapping function $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^3$.

For the feature extraction, N voxels ($X \in \mathbb{R}^3$) were randomly chosen within the FOV of the image, being their relative displacement ($d \in \mathbb{R}^3$) the information to be predicted, i.e., their offset to each vertebra centroid: $d_i = c_i - X_i$. The selection of partial data sets of voxels instead of the whole image series allowed to minimize computational burden. A graphical description of the process is appreciated in Fig. 1.

The problem to identify anatomical structures in CT images is that different human structures may share similar intensity values. Thus, local intensity information might not be sufficiently discriminative. To avoid this limitation, a 3D cuboid [p_x, p_y, p_z] was computed around each randomly selected voxel and divided into blocks of size [b_x, b_y, b_z]

(Fig. 2). Then, from each block, the mean intensity was calculated, having F intensity-based features associated with each training voxel.

The mean intensities over cuboidal regions are computed in a short time using the integral image [17]. The advantage of this technique is that the sum of the voxels over any sub-volume can be calculated in constant time once the integral image over the whole CT volume is obtained, no matter how big the volume is. The integral image is an intermediate representation of an image, where each voxel (x, y, z) is the sum of the voxels immediately adjacent (left, front and up) to x', y', z' in the original image. By definition:

$$II(x, y, z) = \sum_{x' \leq x, y' \leq y, z' \leq z} I(x', y', z')$$

where $I(x', y', z')$ is the original image and $II(x, y, z)$ is the integral image. The mean intensity of any block can be computed as:

$$E[X] = \frac{(II_g - II_e - II_h + II_f) - (II_c - II_a - II_d + II_b)}{N}$$

where $\{a, \dots, h\} \in \mathbb{R}^3$ are the eight vertices of the block and N is the number of voxels within the block.

For the testing stage, once the RRF is trained, given a new unseen image, M voxels ($X' \in \mathbb{R}^3$) were randomly selected on the image, and F intensity-based features ($f'_j \in \mathbb{R}^d$) were extracted in the same way as on the training stage. Through the learned mapping function φ , the predicted displacement was obtained: $d'_j = \varphi(f'_j)$. Knowing the location of the reference voxel and the predicted relative distance vector to the center of a specific vertebral body, the predicted location was computed as $c'_j = d'_j + X'_j$.

From each testing voxel, a predicted location was obtained. Therefore, for each specific vertebral body, M voxels were candidates to be its centroid. The probability of all M voxels to be the vertebra centroid was calculated by obtaining the probability density function of all candidates. This probability aggregation was obtained by using kernel density estimation (KDE). The global maximum of the density function was considered as the predicted location of the vertebral body centroid in the image.

Refinement based on voxel-wise operations

Due to the expected population variability in spine curvatures, a refinement step was added in order to adapt the centroid detection to the patient-specific spine morphology. For this purpose, we performed an image binarization, using a fixed 200 Hounsfield Units (HU) threshold. As the spinal canal is surrounded by cortical bone, the CT volume was dilated using a structuring element of cylindrical shape with a 3 mm radius and 10 mm height. After dilation, a

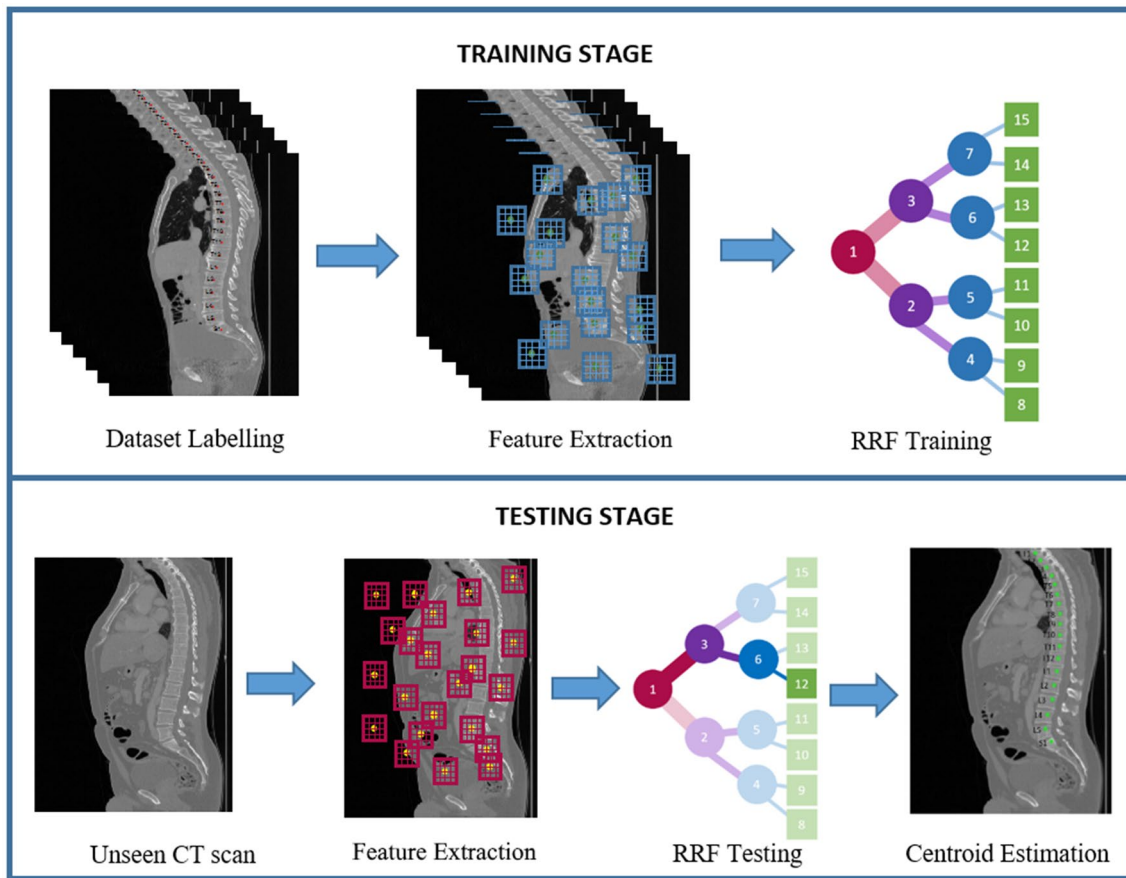


Fig. 1 Flow diagram of the proposed method. Both training (top) and testing (bottom) diagram blocks

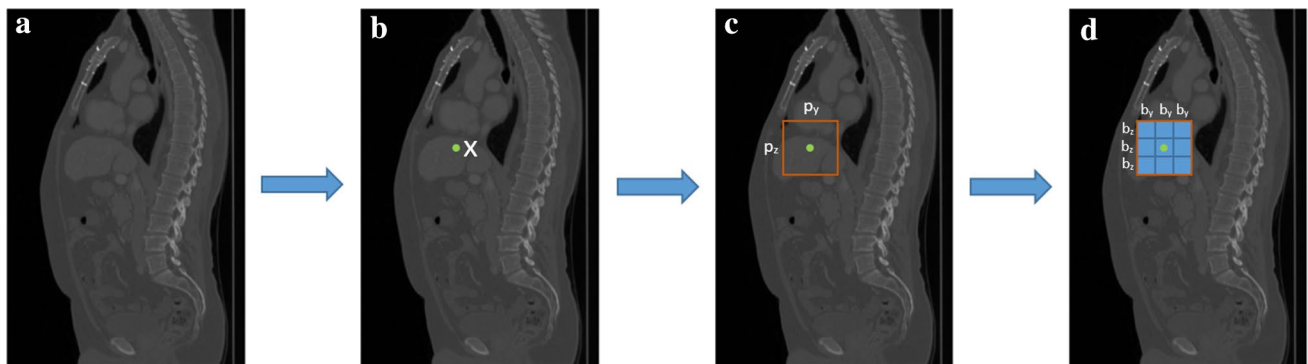


Fig. 2 Workflow from block selection to feature extraction. The x dimension of both the patch (p_x) and blocks (b_x) corresponds to the coronal view. This is an example of how to select the boxes around a selected voxel in an image, where intensity-based features are

extracted. **a** CT volume. **b** Randomly selected voxel. **c** 3D cuboid. **d** Sub-division of the 3D cuboid into blocks. The distance from the selected voxels to a concrete vertebra used to train the forest is also represented

logical NOT operation was performed. At this point, the background was removed and the spinal canal was isolated removing regions with an area lower than 500 mm^3 and adding a boundary condition to detect the spinal canal only in the posterior region of the image. Finally, the spinal canal centerline was extracted in 3D space.

In Fig. 3, the flow diagram for the spinal canal detection is shown.

Once the spinal canal was detected, the obtained curve was displaced 2 cm in the y axis in the posterior-anterior direction of the image, adapting the curve to the centerline of the spine. This displacement was defined after testing

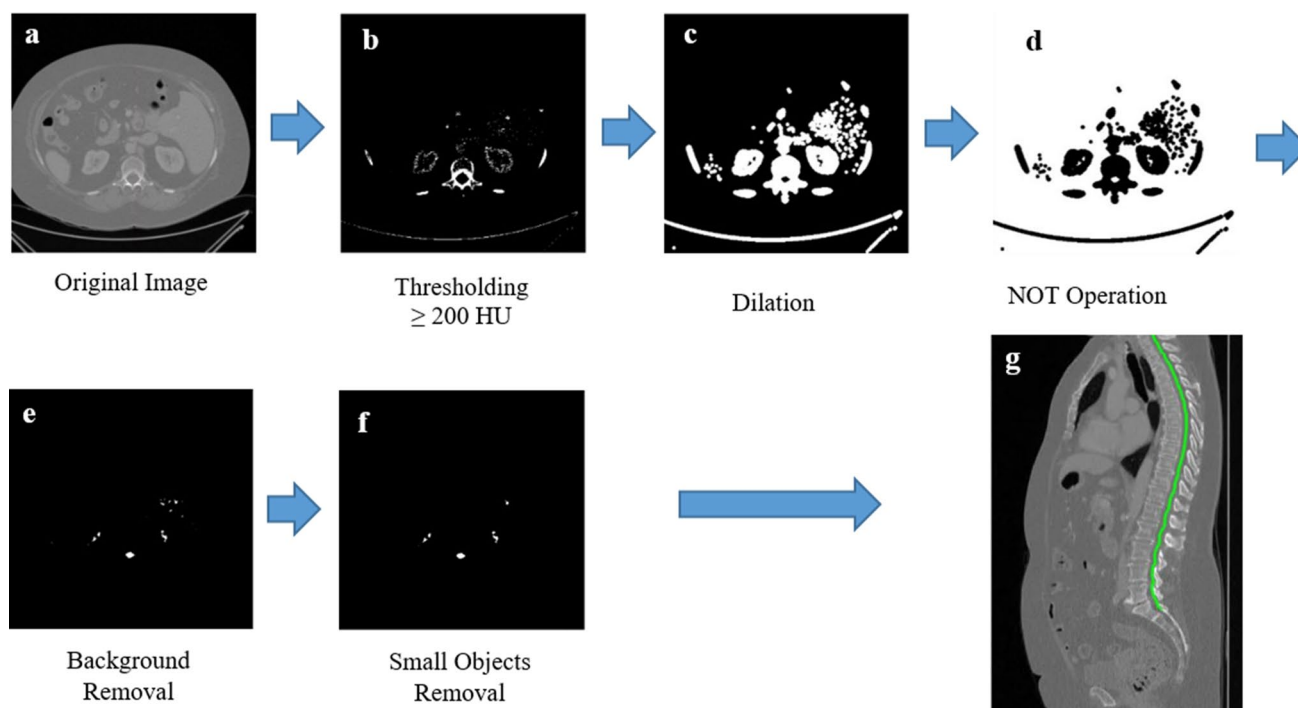


Fig. 3 Refinement step flow diagram. **a** Original image. **b** Image thresholding at 200 HU. **c** Image dilation by applying a cylindrical structuring element. **d** Logical NOT operation. **e** Background removal. **f** Objects smaller than 500 mm^3 removal. **g** Spinal canal centerline

several options, obtaining the best performance adjusting the displacement to 2 cm.

With the spine centerline detection, the previously obtained centroid coordinates (x, y, z) were transformed to the final (x', y', z') coordinates. As a last refinement step, for each $z = z'$ point, its corresponding (x, y) coordinates were changed to (x', y') , adapting each predicted vertebra centroid to the spine curvature.

Training

The parameters used in the training–testing stages can be appreciated in Table 1.

Therefore, considering the whole training dataset, the RRF was trained with 45,824 samples, having 256 features each one. All these features were used to train the RRF, with a total training time of 3 h.

Table 1 Parameters used in the training–testing steps

Parameter	Description	Value
N	Number of aleatory training points	30,000
M	Number of aleatory testing points	50,000
$[p_x, p_y, p_z]$	Patch size (mm)	$40 \times 40 \times 120$
$[b_x, b_y, b_z]$	Block size (mm)	$10 \times 10 \times 30$
F	Number of extracted features from each aleatory-selected point	256
N. trees	Number of trees	20
Max. features	Number of features to consider when looking for the best split	16
Max. depth	Maximum depth of the tree	24
Min. split node	Minimum number of samples required to split an internal node	8

The upper rows show the parameters for the feature extraction step. The lower rows show the parameters used to build the RRF

Testing and performance evaluation

To test a new image, unseen on the training stage, 50.000 random voxels were selected, with 256 features for each testing voxel. Total testing time was 3 min.

To evaluate the performance of the network, the distance between the predicted position of each centroid and the real one, defined by previous expert annotations, as well as the identification rate was calculated. A vertebra was correctly identified if the estimated centroid was within 2 cm of the real one.

Results

An initial detection was performed applying decision forests. Then, the detected centroid position was refined by obtaining the position of the spinal canal (Fig. 4).

In Fig. 5, the localization error on each direction (x , y , z) can be appreciated. For all vertebrae, the median of the distance between the predicted centroid position and the real one is calculated. The minimum error is obtained on the x direction (left–right), and the maximum one is obtained on the z direction (head–feet). This occurs mainly because with the refinement step; the errors obtained on both the x and y (anterior–posterior) directions were minimized.

In Fig. 6, the localization error on each direction per vertebra is detailed. It can be seen that the localization error on x direction is very similar for all vertebrae. However, the localization error for both the y and z directions depends on the corresponding vertebra.

If the distance in all directions is considered, the vertebrae with the minimum and maximum localization errors are easily obtained (Fig. 7). The minimum localization error is at the central thoracic vertebrae (T9–T11), and the maximum localization error is on the upper thoracic vertebrae

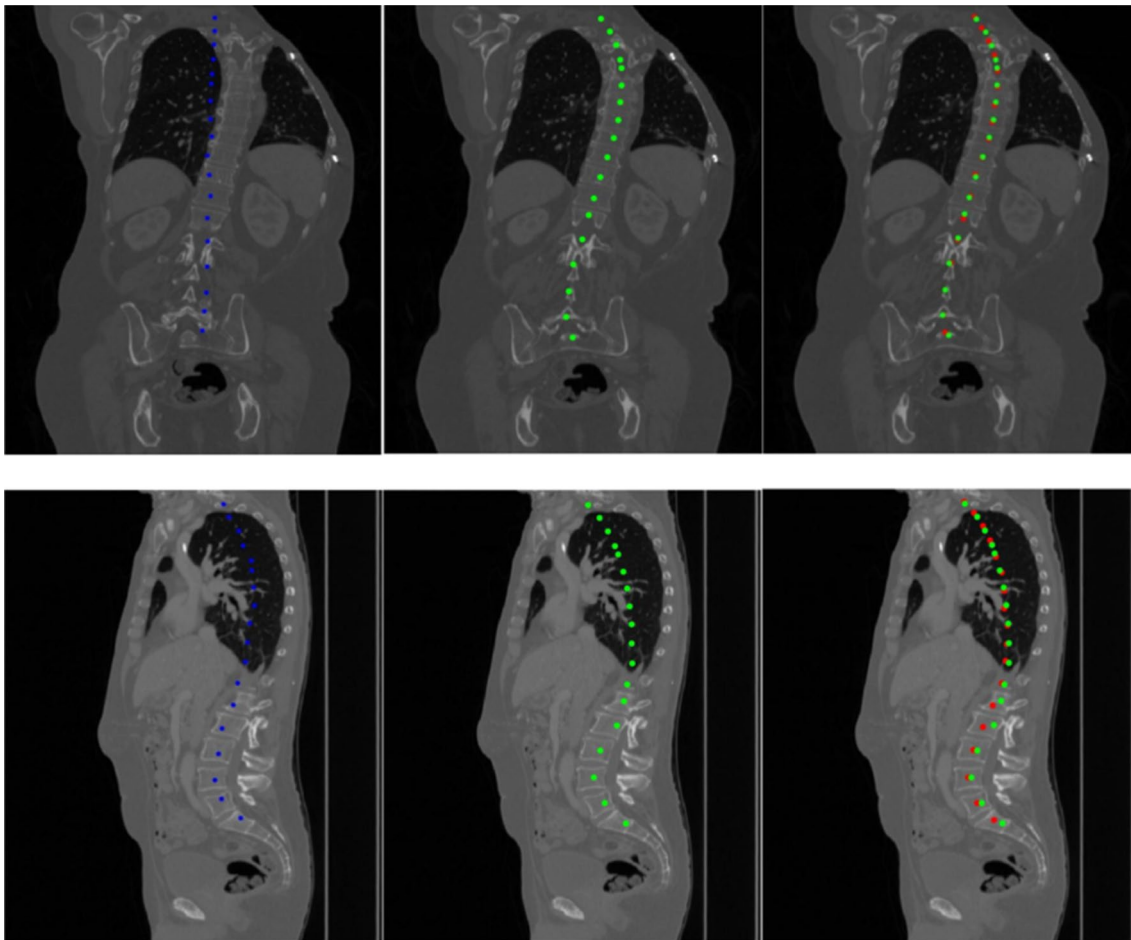


Fig. 4 Vertebral bodies localization after the rough detection by applying decision forests (left) and after the refinement by detecting the spinal cord position (middle). The predicted positions are compared with the annotation of an expert (right). Both coronal (top) and sagittal (bottom) views are shown (blue: rough detection;

green: refinement; red: expert annotation). All centroids are shown in the same slice to provide a 2D visualization of the obtained results, although the real volume is 3D. It is a case of a patient with a significant scoliosis; this is the reason why some vertebrae are not visible on the sagittal view

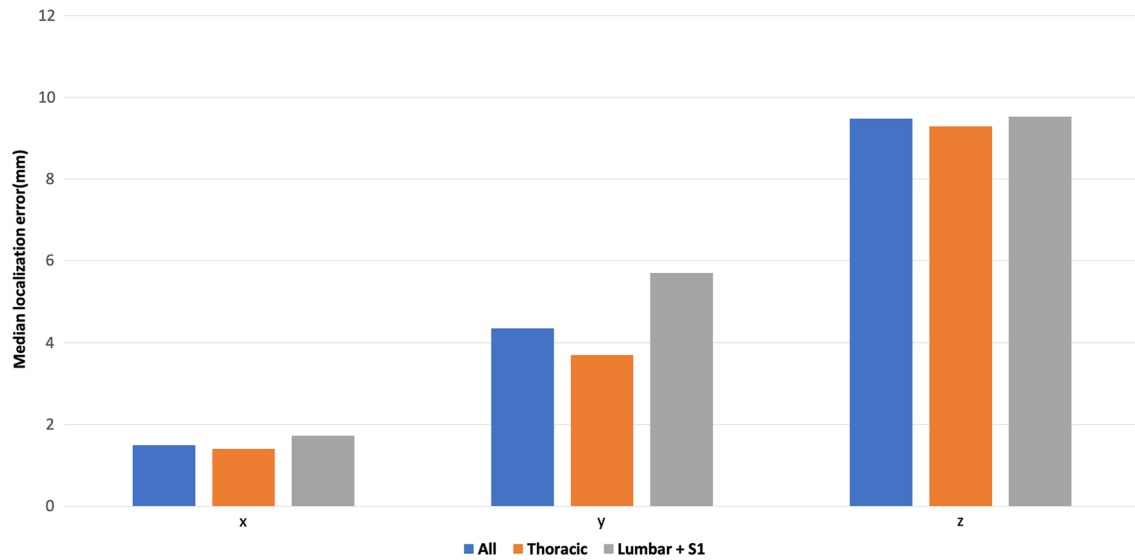


Fig. 5 Median localization error per axis of all vertebrae (blue), the thoracic region (orange) and the lumbar and sacrum region (gray)

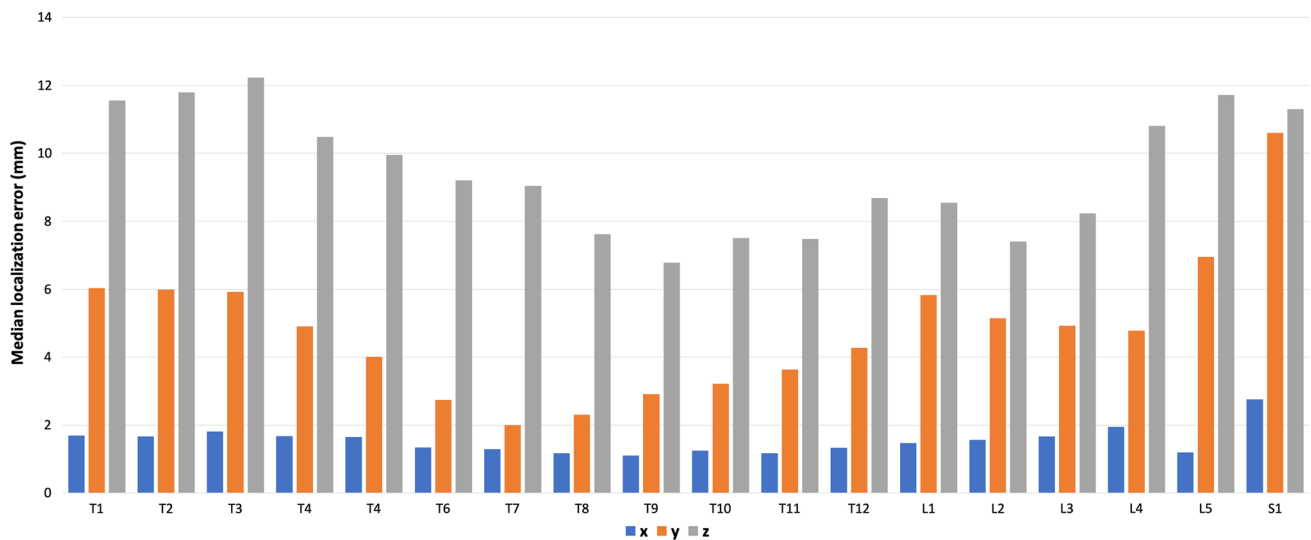


Fig. 6 Median localization error in mm per vertebra and direction

(T1–T4). In the lumbar region, the localization error is very similar for all vertebrae.

The localization error and the identification rate obtained after rough detection and after refinement are summarized in Table 2.

In Table 2, it can be seen the improvement of both the distance between the predicted vertebrae position and the real one and the identification rate after refinement. The mean distance error decreases from 15.7 to 13.7 mm, and the identification rate increases from 72.22 to 77.99%. After the rough detection, the identification rate is similar both in thoracic and lumbar regions; however, after refinement,

this rate increases in both regions, increasing mainly in the thoracic region.

Discussion

In this work, an approach for the automatic localization and identification of the vertebral bodies in CT scans has been proposed using RRF. The algorithm has been tested using a dataset including both healthy and pathological cases and where no assumptions about the visible region have been made, therefore working with arbitrary FOVs.

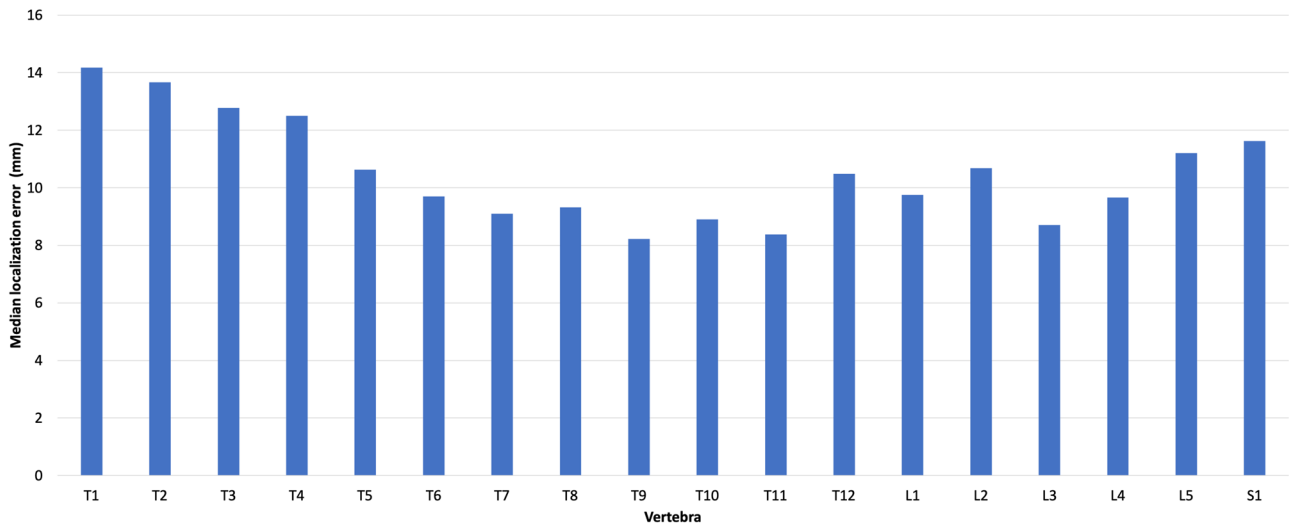


Fig. 7 Median localization error per vertebrae

Table 2 Localization errors in mm obtained after rough detection (left) and after refinement (right)

Region	Rough detection				Refined detection			
	Median	Mean	Std	ID. rate	Median	Mean	Std	ID. rate
All	14.163	15.727	9.073	72.22%	10.335	13.734	10.318	77.99%
Thoracic	14.554	15.889	8.375	72.17%	10.172	13.045	9.478	79.56%
Lumbar + S1	12.836	15.403	10.337	72.33%	10.710	15.112	11.716	74.84%

All the methodologies presented on [10–14] used the same dataset, presented by Glocker et al. in [10], both to train and to test their performances. However, this dataset is built of spine-focused CT scans by using cropped images. Under our point of view, better clinical integration can be achieved by the use of the original images. CT scans are mostly acquired including the whole abdominal area, where, apart from the spine, additional anatomical structures are included. In this way, we gain spatial information; however, the computational burden needed to process these images is higher. To integrate an algorithm into clinical routine, a key aspect is the use of RWD on its development and validation. This is the reason why we decided to use our own dataset, acquired directly from the PACS of a tertiary hospital.

Further improvements to this work are possible. Considering also the cervical region on the training stage to predict the location of these vertebrae in those images where this region is present. In our work, cervical region was not included because, from all the clinical scans collected, only a few of them included the cervical region. These images were not enough to train a RRF with high identification rate on these vertebrae; therefore, they were excluded. Therefore, in our method, cervical vertebrae can be present in the images under study; however, their position will not be predicted.

Conclusion

RRF allows a reliable vertebrae localization and identification in real-world CT data. Due to the high variability in the field of view and anatomical landmarks between different CT scans, it might be very difficult to consistently obtain a high-accuracy prediction of vertebrae position. Therefore, future work will focus on further improving these results combining other AI techniques with decision forests and using more complex features in order to reduce the identification errors obtained in the present work.

Funding Funding was provided by Asociación para la Investigación y el Desarrollo en Resonancia Magnética – ADIRM, Ministry of Economy, Industry and Competitiveness (Grant No. DPI2014-53401-C2-2-R).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Zukić D, Vlasak A, Dukatz T, Egger J, Horinek D, Nimsky C, et al (2012). Segmentation of vertebral bodies in MR images. In: Goesele M, Grosch T, Preim B, Theisel H, Toennies K (eds) Proceeding of 17th international workshop on VMV, pp 135–142. <https://doi.org/10.2312/pe/vmv/vmv12/135-142>
- Egger J, Kapur T, Dukatz T, Kolodziej M, Zukić D, Freisleben B et al (2012) Square-cut: a segmentation algorithm on the basis of a rectangle shape. *PLoS ONE* 7(2):e31064. <https://doi.org/10.1371/journal.pone.0031064>
- Ayed IB, Punithakumar K, Minhas R, Joshi R, Garvin GJ (2012) Vertebral body segmentation in MRI via convex relaxation and distribution matching. In: Proceedings of medical image computing and computer-assisted intervention—MICCAI 2012, pp 520–527. https://doi.org/10.1007/978-3-642-33415-3_64
- Herring J, Dawant B (2001) Automatic lumbar vertebral identification using surface-based registration. *J Biomed Inform* 34(2):74–84. <https://doi.org/10.1006/jbin.2001.1003>
- Ma J, Lu L (2013) Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. *Comput Vis Image Underst* 117(9):1072–1083. <https://doi.org/10.1016/j.cviu.2012.11.016>
- Chu C, Belavý D, Armbrecht G, Bansmann M, Felsenberg D, Zheng G (2015) Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method. *PLoS ONE* 10(11):e0143327. <https://doi.org/10.1371/journal.pone.0143327>
- Chwialkowski M, Shile P, Pfeifer D, Parkey R, Peshock R (1991) Automated localization and identification of lower spinal anatomy in magnetic resonance images. *Comput Biomed Res* 24(2):99–117. [https://doi.org/10.1016/0010-4809\(91\)90023-P](https://doi.org/10.1016/0010-4809(91)90023-P)
- Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C (2009) Automated model-based vertebra detection, identification, and segmentation in CT images. *Med Image Anal* 13(3):471–482. <https://doi.org/10.1016/j.media.2009.02.004>
- Schmidt S, Kappes J, Bergtholdt M, Pekar V, Dries S, Bystrov D, Schnörr C (2007) Spine detection and labeling using a parts-based graphical model. *IPMI* 4584:122–133. https://doi.org/10.1007/978-3-540-73273-0_11
- Glocker B, Feulner J, Criminisi A, Haynor D, Konukoglu E (2012) Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Medical image computing and computer-assisted intervention—MICCAI 2012, pp 590–598. https://doi.org/10.1007/978-3-642-33454-2_73
- Glocker B, Zikic D, Konukoglu E, Haynor D, Criminisi A (2013) Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In: Medical image computing and computer-assisted intervention—MICCAI 2013, pp 262–270. https://doi.org/10.1007/978-3-642-40763-5_33
- Suzani A, Seitel A, Liu Y, Fels S, Rohling R, Abolmaesumi P (2015) Fast automatic vertebrae detection and localization in pathological CT scans—a deep learning approach. In: Lecture notes in computer science, pp 678–686. https://doi.org/10.1007/978-3-319-24574-4_81
- Chen H, Shen C, Qin J, Ni D, Shi L, Cheng J et al. (2015) Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In: Lecture notes in computer science, pp 515–522. https://doi.org/10.1007/978-3-319-24553-9_63
- Yang D, Xiong T, Xu D, Zhou SK, Xu Z, Chen M, Park J, Grbic S, Tran TD, Chin SP, Metaxas D (2017) Deep image-to-image recurrent network with shape basis learning for automatic vertebra labeling in large-scale 3D CT volumes. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 498–506. https://doi.org/10.1007/978-3-319-66179-7_57
- Liao H, Mesfin A, Luo J (2018) Joint vertebrae identification and localization in spinal CT images by combining short- and long-range contextual information. *IEEE Trans Med Imaging* 37(5):1266–1275. <https://doi.org/10.1109/TMI.2018.2798293>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Viola P, Jones M (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154. <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.