Master thesis

MSc Technical Medicine

June 2023

Automatic segmentation of the spinal joints and intervertebral disks on low-dose computed tomography: an atlas-based approach

W.R.P. (Wouter) van der Heijden

Amsterdam UMC
Universitair Medische Centra

TU Delft

AUTOMATIC SEGMENTATION OF THE SPINAL JOINTS AND INTERVERTEBRAL DISKS ON LOW-DOSE COMPUTED TOMOGRAPHY: AN ATLAS-BASED APPROACH

BSc. W.R.P. van der Heijden

Student number : 4550153

27 June 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)
Dept. of Radiology and Nuclear Medicine, Amsterdam UMC, location VUmc
Dept. of Rheumatology, Amsterdam UMC, location VUmc
November 2022 – July 2023

Supervisor(s):

Prof. dr. R. Boellaard	Technical Supervisor
Prof. dr. C.J. van der Laken	Medical Supervisor
Dr. F.H.P. van Velden	Technical Supervisor

Thesis committee members:
Dr. F.H.P. van Velden, LUMC (chair)
Prof. Dr. R. Boellaard, Amsterdam UMC
Prof. Dr. C.J. van der Laken, Amsterdam UMC
Dr. J.P.H.J Rutges, Erasmus MC

Automatic segmentation of the spinal joints and intervertebral disks on low-dose computed tomography: an atlas-based approach

Wouter R.P. van der Heijden[1,2], Conny J. van der Laken[2], Gerben J.C. Zwezerijnen[1], Floris H.P. van Velden[3] and Ronald Boellaard[1]

[1]Amsterdam UMC, Department of Radiology and Nuclear Medicine, De Boelelaan 1117, Amsterdam, Netherlands

[2]Amsterdam UMC, Department of Rheumatology, De Boelelaan 1117, Amsterdam, Netherlands

[3]Leiden University Medical Center, Department of Radiology, Albinusdreef 2, Leiden, Netherlands

My affiliation:

First and corresponding author:

Wouter R.P. van der Heijden

Department of Radiology & Nuclear Medicine

Department of Rheumatology

Amsterdam University Medical Centers, Location VUmc

De Boelelaan 1117

1081 HV Amsterdam, The Netherlands

Email: w.r.p.vanderheijden@amsterdamumc.nl

Abstract word count: 390

Manuscript word count: 3798

Running title: Automatic segmentation of the spinal joints and intervertebral disks on low-dose computed tomography: an atlas-based approach.

Manuscript category: Original Research

Abstract

Introduction: Spondyloarthritis (SpA) belongs to the chronic inflammatory rheumatic diseases, and primarily affects the axial skeleton. Quantitative ^{18}F-NaF PET/CT is a new imaging approach that shows promise for accurate diagnosis and treatment assessment. Manual segmentation of low-dose computed tomography (LDCT) for quantitative feature extraction is time-consuming and subjective, and can be replaced by automatic methods. This study aims to develop and validate an automatic algorithm for the segmentation of the spinal joints and intervertebral disks (IVD's) on LDCT using two different approaches.

Methods: Two methods for spinal structure segmentation were developed and compared. Both methods used segmentations of bony structures obtained from the TotalSegmentator algorithm. The first method employed morphological dilation and erosion operations to localise the joints and IVD's, while the second method used a multi-atlas-based method approach with partial atlases and corresponding manually segmented labelmaps. The performance of the methods was assessed on ten manually segmented LDCT's using sensitivity, and maximum and average Hausdorff distance (HD) for IVD's and the sacroiliac joints (SIJ) and mean error distance for the smaller joints. The reproducibility of the methods was evaluated using a set of 20 LDCT test-retest images.

Results: The atlas-based method achieved significantly better maximum HD (8.45 (1.80) vs. 9.64 (5.83) ($p = 0.002$)) and sensitivity (0.79 (0.22) vs. 0.61 (0.30) ($p < 0.001$)) for all IVD's combined compared to the morphological method. The atlas-based method also outperformed the morphological method for the facet joints, costovertebral joints and costotransverse joints with a mean error distance of 4.71 mm (2.72) vs 6.90 mm (4.80) ($p < 0.001$). For the thoracic IVD's the morphological method showed significantly better average HD (1.48 (1.03) vs. 1.72 (0.53) ($p = 0.018$)) and maximum HD (6.97 (3.36) vs. 8.22 (1.66) ($p < 0.001$)) than the atlas-based method. In the reproducibility assessment on the test-retest scans, the atlas-based method outperformed the morphological method for all metrics and structures, with average HD's well below the voxel resolution (< 2 mm).

Conclusion: We present the first methods for automatic segmentation of the spinal structures on LDCT. The atlas-based method seems to be the most suitable algorithm, achieving average HD's below the voxel size, and maximum errors below one centimetre. However, it is dependent on accurate segmentation by the TotalSegmentator algorithm. Further research is warranted to investigate the influence of the segmentation results on the extraction of quantitative PET information.

Keywords: Low-Dose Computed Tomography, Automatic Segmentation, Spinal joints, Intervertebral Disks, Spondyloarthritis

1. Introduction

Spondyloarthritis (SpA) such as ankylosing spondylitis (AS) and psoriatic arthritis (PsA) belongs to a group of chronic inflammatory rheumatic diseases that primarily affect the axial skeleton and peripheral entheses, leading to significant morbidity and impairment of the quality of life (1). However, diagnosis of SpA can be a time-consuming process, taking years to establish an accurate diagnosis (2). Early diagnosis of SpA is crucial as it enables an earlier initiation of effective treatment, ultimately improving patient outcomes (3). In addition, early efficacy assessment of treatment is needed to prevent prolongation of ineffective treatment with potential side-effects and increased risk of progression of structural damage.

Positron emission tomography (PET), combined with computed tomography (CT) is a powerful diagnostic and monitoring tool widely used for the detection of malignant and benign diseases including inflammatory pathologies (4-6). By providing information on tissue molecular activity and characteristics, PET imaging plays a crucial role in assessing disease activity, treatment response and patient prognosis. In the case of axial SpA, ^{18}F-NaF PET/CT can effectively detect molecular bone formation activity that matches typical disease locations which are known for SpA, anatomically corresponding to e.g. enthesopathy, syndesmophytes and sacroiliitis (7, 8). Moreover, In PsA patients, ^{18}F-NaF-PET-avid lesions are commonly observed in the cervical facet joints, as well as the anterior thoracic and lumbar vertebrae (9, 10).

Although visual interpretation of PET images by nuclear medicine specialists is informative, it is time-consuming, subjective and prone to inter- and intra-observer variability (11-13). Consequently, there is a growing interest in developing automatic methods that can enhance efficiency and consistency in the quantitative analysis of spinal PET-images (14-19). Extracting various quantitative parameters from PET images, such as standardized uptake value (SUV), total lesion glycolysis (TLG), and texture or histogram features, provides valuable insights into disease characteristics, heterogeneity and treatment response, enabling clinicians to make more informed decisions (20). For automatic extraction of the quantitative PET features within a specific anatomical structure, segmentation of the low-dose CT (LDCT) scan acquired concurrently with the PET is required (21, 22). The LDCT scan offers more anatomical information than the PET image, and can be employed for more reliable segmentation.

While numerous algorithms for automatic segmentation of the bony spinal structures on LDCT have been described (23-26), automatic segmentation of the spinal joints and intervertebral disks (IVD's) remains a relatively new research area. Piri *et al.* developed a method for segmenting lumbar joints and IVD's on diagnostic CT scans of patients with lower back pain (27). Their approach utilized a pre-existent UNet-based (28) segmentation network RECOMIA (24) for initial segmentation of the vertebrae, followed by morphological dilation and erosion operations to segment the lumbar IVD's and facet joints (FJ's). However, there are currently no algorithms for automatic segmentation of all spinal structures on LDCT images.

In this paper, we present and compare two automatic methods for segmenting IVD's and spinal joints on LDCT scans. Our primary objective is to develop robust algorithms that can accurately segment key anatomical components, including the IVD's, FJ's, costovertebral joints (CVJ), costotransverse joints (CTJ), and sacroiliac joints (SIJ) (29). The optimal method should be accurate, robust, more time-efficient than manual segmentation and require minimal user interaction.

2. Methods

Data acquisition

^{18}F-NaF PET/CT images of PsA and AS patients were retrospectively analysed. Patients underwent ^{18}F-NaF PET/CT if they met (in the case of PsA) the Classification criteria for Psoriatic arthritis (CASPAR) (30) or were diagnosed with PsA by the treating rheumatologist, and had enthesis of at least one site according to the Maastricht Ankylosing Spondylitis Enthesitis Score (MASES)(31) and/or the Spondyloarthritis Research Consortium of Canada (SPARCC) enthesis index (32). AS patients were included if they fulfilled the modified New York criteria (33) and had a high disease activity with a BASDAI (34) score greater than or equal to 4. More information about the patient group has been described in previous publications (8-10). PET/LDCT images were acquired using either Gemini TF or Ingenuity TF (Philips Healthcare, Best, The Netherlands), or Biograph mCT Flow VG70A (Siemens Healthineers, Erlangen, Germany) PET/CT-scanners. Patients received a mean radioactivity dose of 101 MBq, with a standard deviation of 4 MBq. 45 minutes after dose injection whole-body scans were acquired. The PET scan was preceded by a LDCT (30 mAs, 120 kVp). The PET images were normalised and corrected for attenuation using the LDCT. All scans were reconstructed to a resolution of 4 x 4 x 4 mm^3.

Image Processing

The segmentation of the bony structures of the spine was performed using the TotalSegmentator (25) network, which is based on the nn-UNet architecture (35). This algorithm had been trained and validated for the segmentation of 104 structures on diagnostic CT and LDCT scans. The relevant structures for joint segmentation included individual vertebrae, ribs, sacrum and pelvis. These structures where extracted from the original segmentation of the 104 structures.

The structures that were segmented in this study included 24 IVD's, 48 FJ's, 44 CVJ's, 20 CTJ's. and 2 SIJ's. Since the main application for the joint segmentation algorithm would be quantification of PET images, and the exact determination of the VOI depends on the specific clinical question, the FJ's, CVJ's and CTJ's were not segmented as volumes, but as one point representing the centre of the joint. To create an appropriate VOI, a spherical dilation operation could be applied to this point. Two methods were developed for the segmentation of the spinal structures on LDCT and compared using manual segmentations as gold standard.

Morphological operations

For the first method, the segmented vertebrae were separated into an anterior and two posterior parts using Matlab R2017b with a variation of the method developed by Piri *et al* (27). The vertebral body was separated from the transverse processes by applying a 8 mm radius spherical erosion operation to the initial bone segmentation. The largest remaining component was dilated by 12 mm, to obtain the segmentation of the vertebral body (VB). The dilation radius was larger than the erosion radius to ensure accurate segmentation of all voxels of the anterior vertebra. Voxels that were not initially included in the bone segmentation were removed from the dilated volume. The remaining part of the initial segmentation was classified as the posterior vertebra. The posterior vertebrae were further divided into left and right parts, using the center of mass (COM) of the original vertebra segmentation. Voxels to the left of the sagittal plane of the COM were classified as the left posterior vertebra (LPV), while voxels to the right were classified as the right posterior vertebra (RPV) This division resulted in

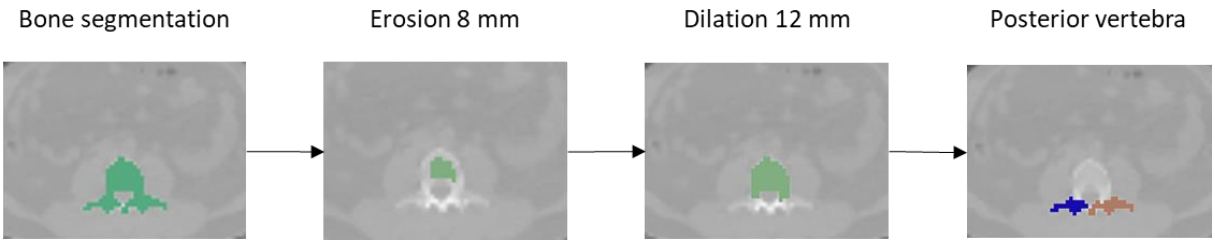

Figure 1: Vertebra Segmentation. Using morphological operations the vertebra was divided into three parts: The vertebral body (green), the right posterior vertebra (blue) and the left posterior vertebra (brown).

three parts for each vertebra (Figure 1). The IVD's and joints were segmented using these three bone segmentations.

For each IVD, the two adjacent VB's were progressively dilated with a 4 mm radius spherical structuring element, until an overlap between the bodies occured. The overlapping dilated voxels were classified as IVD. The same dilation procedure was applied to the LPV and RPV for the FJ segmentation, as well as to the ribs for the CVJ's and CTJ's. To precisely locate the small joints (all joints except for the SIJ's), a single voxel was chosen to be the center of the joint. This location was determined by calculating the COM of the overlapping dilated voxels. This approach ensured accurate localization of the joint centers for further analysis and quantification.

Atlas-based

The atlas-based method involved manually segmenting five LDCT scans to be used as atlases. Co-registration of the atlases with the target scan was performed to obtain spinal structure segmentations, as will be explained later. For appropriate atlas selection the distance between the COM's of C1 and the sacrum in the bone segmentations was calculated for all available LDCT scans, providing an estimate of the spinal length. To ensure diversity, the LDCT scans of the patients that represented the 10^{th}, 30^{th}, 50^{th}, 70^{th}, and 90^{th} percentile of the spinal length were used as atlases. These images were further processed using 3D slicer v5.0.2 (36) and Matlab R2017b. To optimise co-registration of all structures, the atlases were divided into smaller volumes containing sets of adjacent vertebrae. These volumes were extracted from the bone segmentations using the minimum bounding box of two adjacent vertebrae. The same bounding boxes were applied to the manually segmented labelmaps, resulting in 26 partial atlases and labelmaps (figure 2).

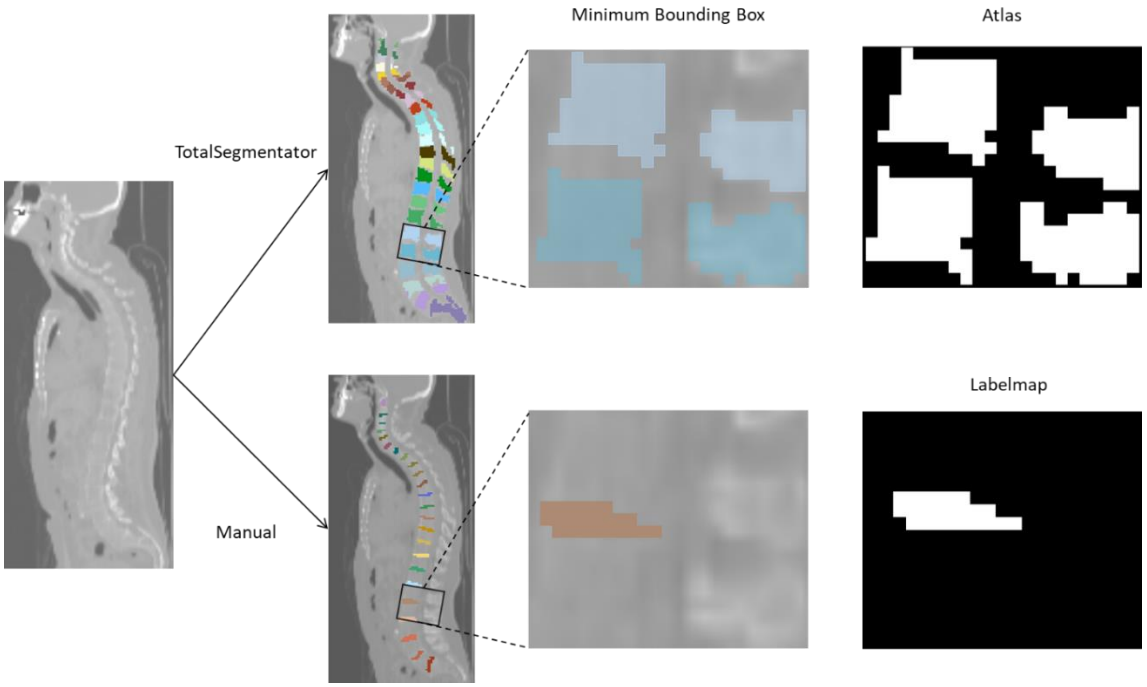

Figure 2: Atlas creation. The atlas LDCT was segmented using TotalSegmentator for the vertebrae, and manually for the spinal joints and intervertebral disks. The minimum bounding box for each set of adjacent vertebrae was extracted from the bone segmentation and saved as atlas. The same bounding box was applied to the manual segmentation, and saved as the labelmap.

Co-registration

To segment the structures on a target LDCT scan, the bone segmentation was obtained using TotalSegmentator. This segmentation was cropped into bounding boxes containing combinations of adjacent vertebrae, similar to the creation of the atlases. To ensure optimal co-registration, a volume of ten voxels with a value of 0 was added around the edges of these cropped images. For each vertebral level, the five corresponding atlases were co-registered onto the target images using the Elastix toolbox (37) (Figure 3). A similarity transform was applied to the atlases, optimized through 100 iterations using stochastic gradient descent as the optimizer and advanced mean squares as the metric. The step size for each iteration was determined adaptively, with a maximum step size of 0.1. The final transform of each atlas was applied to the corresponding labelmap. Majority voting was employed for IVD and SIJ segmentation, assigning a label to a voxel if at least three of the atlases agreed on that label. For the localization of FJ, CVJ, and CTJ, the mean coordinates of the five estimated locations were calculated and classified accordingly (Figure 4). The bounding boxes containing the final segmentations were integrated back into the target image.

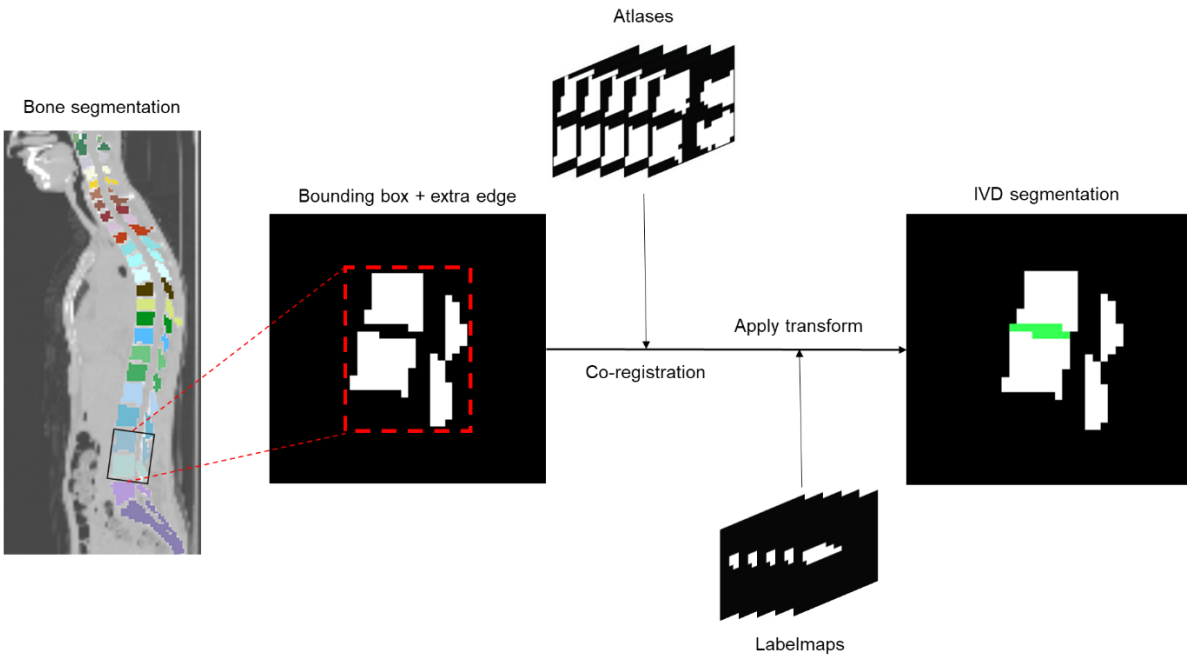

Figure 3: Atlas-based segmentation. The bone segmentation of the target LDCT was cropped, resulting in binary images of each set of two adjacent vertebrae. To improve co-registration performance an extra edge was added around the mask. Five atlases were co-registered to the target images, after which the same transformation was applied to the corresponding labelmaps. The final segmentation was determined using majority voting for intervertebral disks and sacroiliac joints, and the mean location for the facet joints and rib joints.

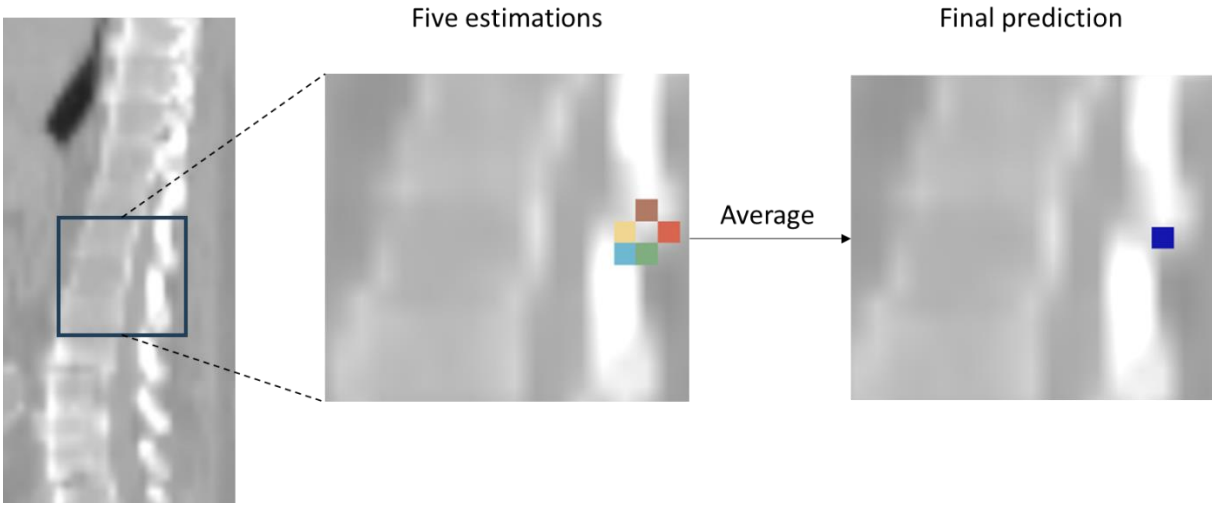

Figure 4: Facet joint localisation. For the segmentation of the facet joints and rib joints the mean estimated location of the five atlases was calculated.

Validation

The validation of the segmentation methods involved several assessments to evaluate the performance and reproducibility. To assess the performance, ten random LDCT scans from the SpA cohort were automatically segmented using both automatic methods, and manually segmented for ground truth definition. The evaluation metrics included the maximum and average Hausdorff distance (HD) and sensitivity for the IVD's and SIJ's, as well as the distance between the estimated point and the manually segmented point for the FJ's and rib joints.

The HD metric was chosen due to the small volume of the segmentations. In contrast to other metrics that only consider the overlap between segmentation and ground truth, the HD calculates the closest distance between each point in the segmentation and the ground truth. The maximum HD (Equation 1) represents the largest distance between two sets of points, in this case the segmentation (seg) and ground truth (tr) (38).

$$HD_{max}(seg, tr) = \max\big(d(seg, tr), d(tr, seg)\big) \;\; with\; d(seg, tr) = \max_{a \in seg}\big(\min_{b \in tr}|a - b|\big)$$

Equation 1: Maximum Hausdorff distance

Since the maximum HD can be sensitive to outliers, the average HD was calculated as well. The average HD is defined as the average shortest distance of each point in the segmentation to the ground truth, and the average shortest distance of each point in the ground truth to the segmentation (Equation 2) (39).

$$HD_{avg}(seg, tr) = \left(\frac{1}{seg}\sum_{a \in seg}\min_{b \in tr} d(a, b) + \frac{1}{tr}\sum_{b \in tr}\min_{a \in seg} d(b, a)\right)/2$$

Equation 2: Average Hausdorff distance

Reproducibility

To asses reproducibility, the segmentation methods were applied on 20 test-retest LDCT scans of ten lung cancer patients with no spinal metastases (40). These patients underwent ^{18}F-FDG PET/LDCT scans twice within a 3 day time interval. The spine segmentations from the second scan were co-registered onto the spine from the first scan using a rigid transformation. The same transformation was applied to the labelmap, and the similarity between the results was assessed using maximum and average HD, and mean error distance.

Statistical analysis

The performance metrics of both methods were compared using a two-tailed paired sample t-test. Differences in performance were considered significant when the p-value was lower than 0.05. The metrics were calculated for all IVD's and all small joints combined, as well as for each individual group separately. This analysis helped determine the differences in performance between the two segmentation methods.

3. Results

Thirty-eight baseline (prior to initiation or switch of treatment) scans of SpA patients, who underwent ^{18}F-NaF PET/CT, were analysed. All images underwent segmentation by the TotalSegmentator algorithm. The mean vertical distance between the COM of the sacrum the COM of C1 was 58.07 cm with a standard deviation (SD) of 3.77 cm. All lengths were sorted from small to large, and the lengths at the 5th, 13th, 20th, 27th, and 34th indexes were selected as the atlases.

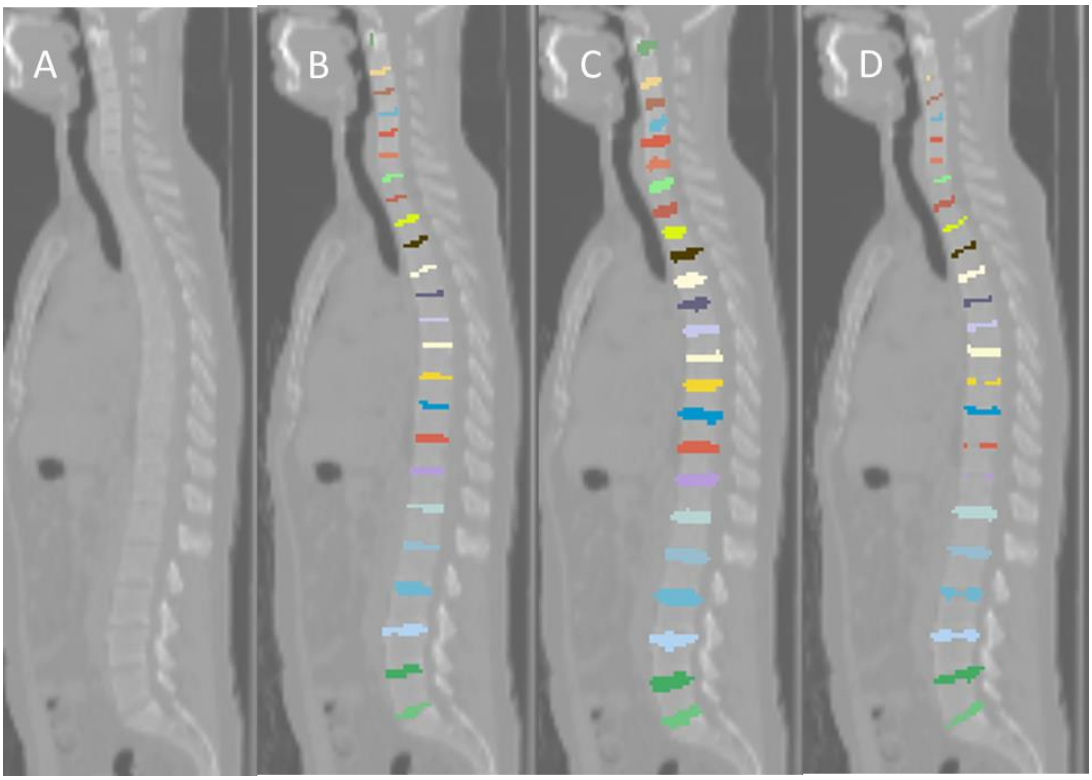

Figure 5: Segmentation results. A: original LDCT. B: Manual segmentation. C: Atlas-based segmentation. D: Morphological segmentation.

The results of the IVD segmentation are visualized in figure 5. In the atlas-based method, the estimated IVD volumes are too large in most cases (Figure 5c), whereas the morphological method (Figure 5d) leads to smaller volumes compared to the manual segmentation (Figure 5b). The segmentation performance for the IVD's and SIJ's is presented in table 1 and figure 6. The atlas-based method (method A) achieved an average HD of 1.98 (SD 1.00) for all IVD's, compared to 2.24 (SD 2.28) for the morphological method (method B). However, the difference was not statistically significant ($p = 0.086$). Method A demonstrated superior performance for all IVD's when the maximum HD (8.45 (1.80) vs 9.64 (5.83), $p = 0.002$) and the sensitivity (0.79 (0.22) vs 0.61 (0.30), $p < 0.001$) were considered. For individual groups, method A showed better performance than method B in the cervical IVD's, lumbar IVD's and SIJ's. However, for thoracic IVD's Method B achieved a significantly better average HD (1.48 (1.03) vs 1.72 (0.53), $p = 0.018$) and maximum HD (6.97 (3.36) vs 8.22 (1.66), $p < 0.001$) than method A. The morphological method showed higher SD's in the average and maximum HD's compared to the atlas-based method, indicating greater variability in segmentation performance between patients and structures. The atlas-based segmentation reached significantly higher sensitivity for all structures, compared to the morphological method, particularly for the SIJ's (0.75 (0.13) vs 0.36 (0.09), $p<0.001$) and the cervical IVD's (0.59 (0.26) vs 0.24 (0.14), $p < 0.001$).

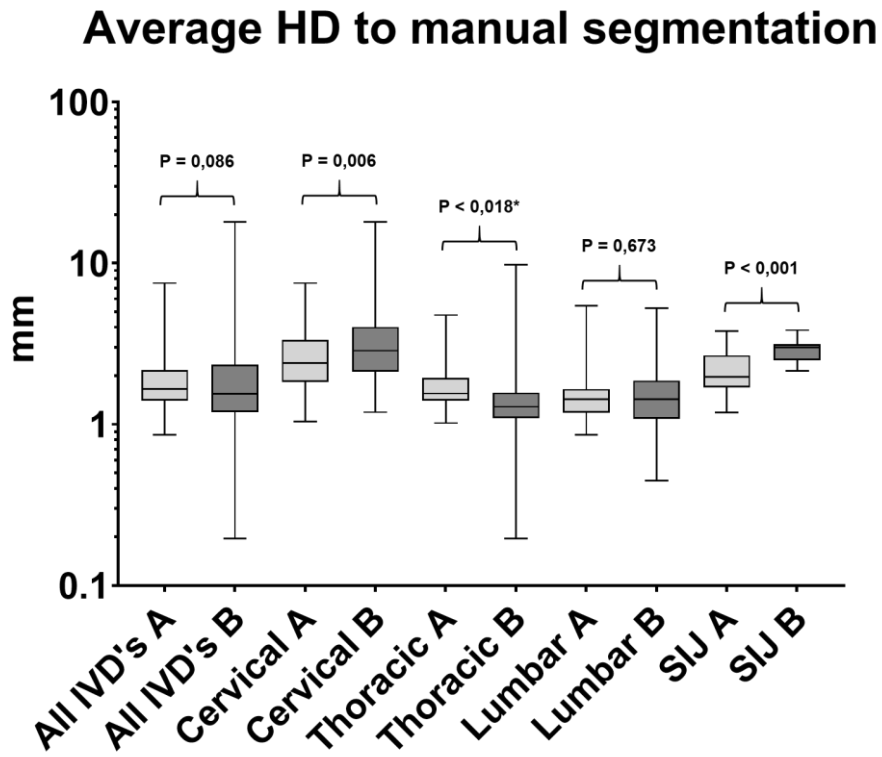

*Figure 6: Boxplots showing the average HD for the intervertebral disk and sacroiliac joint segmentation compared to the manual segmented ground truth. Method A is the atlas-based method and Method B is the morphological method. HD: Hausdorff distance, IVD: Intervertebral disk, SIJ: Sacroiliac joint. *Better performance for method B compared to method A.*

	Average HD (SD) (mm)			Maximum HD (SD) (mm)			Sensitivity (SD)		
	Method A	Method B	p	Method A	Method B	p	Method A	Method B	p
All IVD's	1.98 (1.00)	2.24 (2.28)	0.086	8.45 (1.80)	9.64 (5.83)	0.002	0.79 (0.22)	0.61 (0.30)	<0.001
Cervical	2.74 (1.33)	4.00 (3.32)	0.006	8.38 (1.94)	12.46 (6.03)	<0.001	0.59 (0.26)	0.24 (0.14)	<0.001
Thoracic	1.72 (0.53)	1.48 (1.03)	0.018	8.22 (1.66)	6.97 (3.36)	<0.001	0.86 (0.14)	0.74 (0.20)	<0.001
Lumbar	1.53 (0.68)	1.60 (0.59)	0.673	9.09 (1.75)	12.12 (7.13)	0.003	0.90 (0.13)	0.84 (0.16)	0.025
SIJ	2.14 (0.66)	2.90 (0.45)	<0.001	12.59 (2.35)	13.64 (4.03)	0.315	0.75 (0.13)	0.36 (0.09)	<0.001

Table 1: Performance metrics for intervertebral disk and sacroiliac joint segmentation quality compared to manual segmented ground truth. Method A is the atlas-based method and Method B is the morphological method. HD: Hausdorff distance, IVD: Intervertebral disk, SIJ: Sacroiliac joint.

For FJ's, CVT's and CVR's the mean error distance between the estimated centre and the manually localised centre of the joints were calculated (Figure 7). The atlas-based method consistently outperformed the morphological method for all joints combined (4.71 (2.72) vs 6.90 (4.80), p <0.001) as well as for the individual joint types, with a mean error close to the image resolution of 4 x 4 x 4 mm^3. Similar to the IVD segmentation, the morphological method exhibited higher SD's suggesting more variability in segmentation performance.

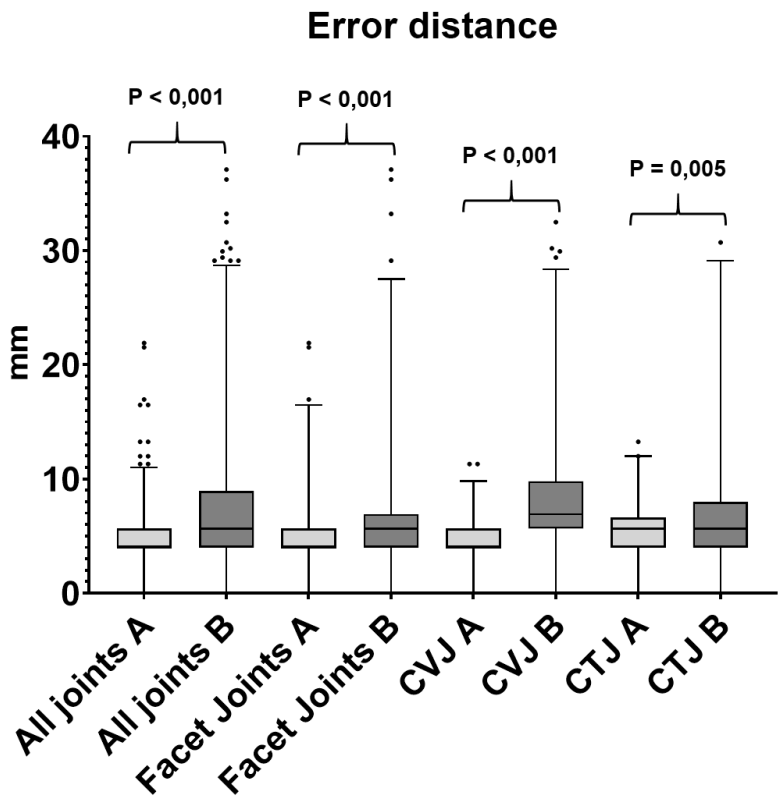

Figure 7: Boxplots showing the error distance between the location estimated by both automatic methods and the manually localised ground truth. The top 1% of all data points are individually visualized. Method A: atlas-based method. Method B: Morphological method. CVJ: Costovertebral joints, CTJ Costotransverse joints.

Reproducibility

Two out of the ten test-retest LDCT pairs were excluded because of substantial differences in spinal shape between the LDCT acquired on day 1 and day 2. In the day 2 scans the upper cervical spines were bent in a different way, possibly as a result of different head positions. Due to this difference no correct alignment could be achieved with rigid co-registration, invalidating the performance metrics. For the remaining 16 scans, both methods were used for segmentation, and the reproducibility metrics are presented in table 2 and figure 8. The atlas-based method achieved significantly better performance for all structures and metrics. The average HD for all IVD's combined was 1.06 (0.76) for the atlas-based method compared to 2.37 (2.43) for the morphological method ($p < 0.001$). For the combination of the FJ's, CVJ's and CTJ's the atlas-based method yielded a mean error distance of 3.87 (2.79), which is below the voxel spacing. In contrast, the morphological method showed poorer performance with a mean error distance of 5.63 (4.13) ($p < 0.001$) (Figure 9).

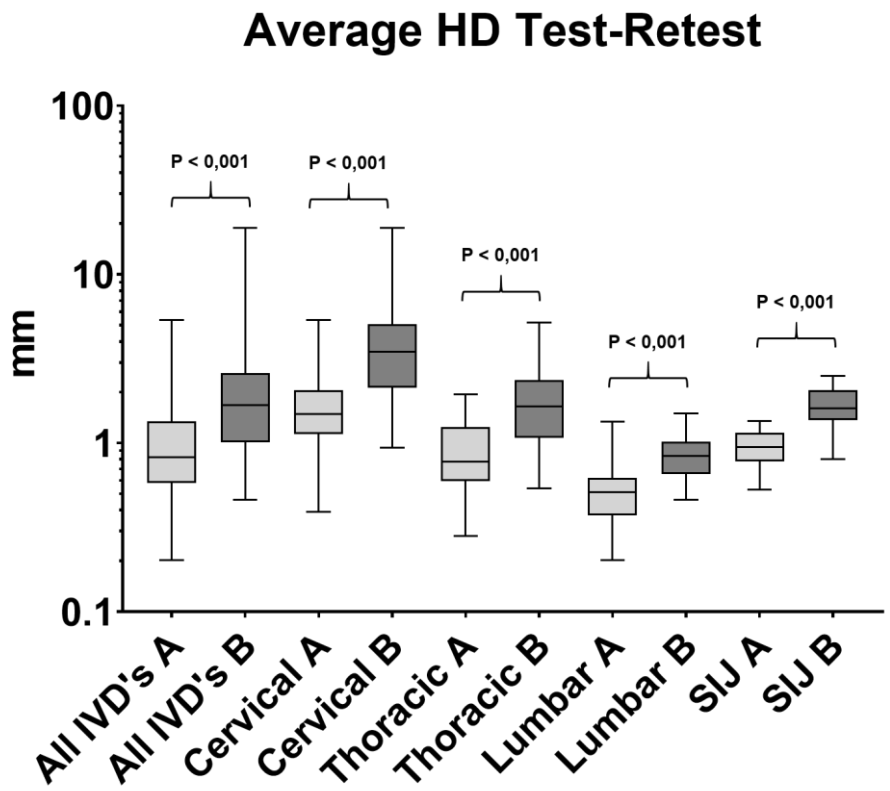

Figure 8: Boxplots showing the average HD for the intervertebral disk and sacroiliac joint segmentation in the test-retest group.. Method A is the atlas-based method and Method B is the morphological method. HD: Hausdorff distance, IVD: Intervertebral disk, SIJ: Sacroiliac joint.

	Average HD (SD) (mm)			Maximum HD (SD) (mm)		
	Method A	Method B	p	Method A	Method B	p
All IVD's	1.06 (0.76)	2.37 (2.43)	<0.001	5.94 (1.68)	8.27 (4.15)	<0.001
Cervical	1.73 (1.00)	4.45 (3.52)	<0.001	6.76 (2.23)	10.50 (5.93)	<0.001
Thoracic	0.91 (0.40)	1.79 (0.89)	<0.001	5.72 (1.30)	7.51 (2.94)	<0.001
Lumbar	0.53 (0.22)	0.86 (0.26)	<0.001	5.36 (1.11)	6.97 (1.61)	<0.001
SIJ	0.96 (0.25)	1.70 (0.45)	<0.001	7.22 (1.68)	9.79 (2.87)	0.015

Table 2: Performance metrics for joint segmentation quality on test-retest LDCT scans that were acquired with a time interval of one day. Method A is the atlas-based method and Method B is the morphological method. HD: Hausdorff distance, IVD: Intervertebral disk, SIJ: Sacroiliac joint.

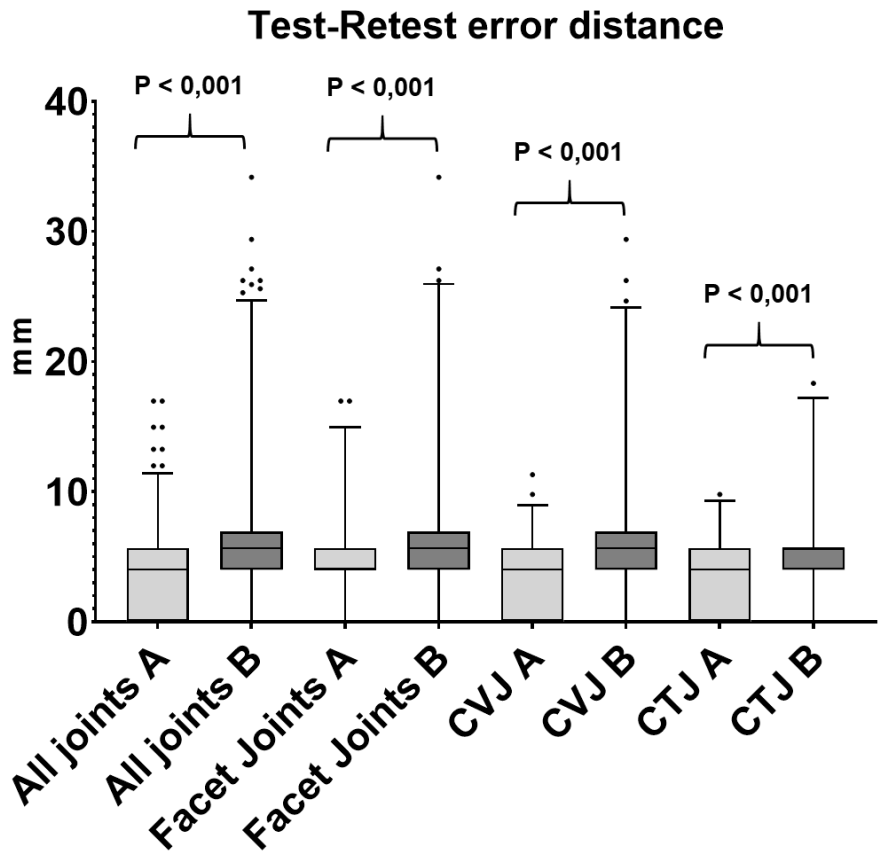

Figure 9: Boxplot of the distance between the location estimated by both automatic methods on the test and retest LDCT. The top 1% of all data points are individually visualized. Method A: atlas-based method. Method B: Morphological method. CVJ: Costovertebral joints, CTJ: Costotransverse joints.

4. Discussion

The atlas-based method demonstrated superior performance compared to the morphological method in most cases when compared with manual segmented ground truths. It was particularly accurate in localizing the FJ's, CVJ's and CTJ's, and achieved higher sensitivities for the IVD's and SIJ's. The atlas-based method also showed superior reproducibility for all structures and metrics.

However, in the segmentation of the thoracic IVD's, the morphological method exhibited significantly better HD metrics compared to the atlas-based method. This discrepancy can be attributed to the spherical dilation operation that is applied to the segmented vertebra. The fixed radius of the dilation, which is a multiple of 4 mm due to the image resolution, roughly approximates the thoracic IVD height. The dilation radius may be too large for cervical IVD's and require additional dilations for lumbar IVD's, leading to larger errors. The segmentation performance for cervical and lumbar IVD's may improve with higher-resolution CT scans where the dilation radius can be chosen more freely.

In most cases the estimated IVD segmentation by the atlas-based method is larger than the manual segmentation. Considering ^{18}F-NaF PET quantification in SpA patients, a small oversegmentation of the IVD's might be beneficial for the reliability. Since specific ^{18}F-NaF uptake commonly occurs around the edges of the IVD, the oversegmentation around the ground truth IVD's could help assessing these lesions. In other PET tracers or pathologies, where more specificity is required, the algorithm could be modified for better results. If the PET quantification should only be performed outside the bony spinal

structures, the bone segmentation can be subtracted from the estimated IVD segmentation. However, in certain cases where there is no space between two vertebrae, which especially might occur in the cervical spine, this will lead to VOIs with extremely small volumes. Another approach to reduce the volume is to modify the majority voting algorithm, requiring a higher number of predictions to assign a certain label to a voxel. The optimal volume for the spherical VOI of the joints should also be subject to further analysis.

One major issue with the atlas-based method is the inaccuracy in bone segmentations, which affects the co-registration performance and structure segmentation. The method relies on the initial bone segmentation obtained from the TotalSegmentator algorithm. Errors in the bone segmentation result in poor performance in the segmentation of the other structures. Manual correction of the bone segmentation is a solution to improve accuracy, but reduces the time-effectiveness and reproducibility of the method. Another approach to reduce dependency on the bone segmentation is to train a convolutional neural network (CNN) specifically for spinal IVD and joint segmentation. However, this requires more training data, initially segmented by the atlas-based method, followed by manual correction.

The LDCT data that were used in this study to test the performance on the manually segmented scans were acquired in SpA patients, which can introduce challenges in co-registration when pathological vertebrae in the atlases are aligned with healthy vertebrae in the test group, and vice versa. For better performance the atlases should be created using CT's of patients with healthy spines. This will lead to easier manual segmentation, which can be rescaled to the desired resolution.

5. Conclusion

To the best of our knowledge, we present the first automatic methods for segmentation of all spinal IVD's and joints on LDCT. The atlas-based method emerges as the most promising approach, exhibiting lower mean error distances, better maximum and average HD performance and higher sensitivity compared to the morphological method. The automatic segmentation algorithms have various potential applications, such as texture analysis on LDCT and diagnostic CT as well as automatic PET image quantification. Further research is needed after the influence of both segmentation methods on automatic PET quantification.

References

1. Dougados M, Baeten D. Spondyloarthritis. Lancet. 2011;377(9783):2127-37.
2. Mease PJ. Suspecting and Diagnosing the Patient with Spondyloarthritis and What to Expect from Therapy. Rheum Dis Clin North Am. 2022;48(2):507-21.
3. McGonagle D, Hermann KGA, Tan AL. Differentiation between osteoarthritis and psoriatic arthritis: Implications for pathogenesis and treatment in the biologic therapy era. Rheumatology (United Kingdom): Oxford University Press; 2015. p. 29-38.
4. Patel PY, Dalal I, Griffith B. [18F]FDG-PET Evaluation of Spinal Pathology in Patients in Oncology: Pearls and Pitfalls for the Neuroradiologist. American Journal of Neuroradiology: American Society of Neuroradiology; 2022. p. 332-40.
5. Pean De Ponfilly – Sotier M, Seror R, Nocturne G, Besson FL. 18F-FDG PET molecular imaging: A relevant tool to investigate chronic inflammatory rheumatisms in clinical practice? Frontiers in Medicine: Frontiers Media S.A.; 2022.
6. Al-Zaghal A, Ayubcha C, Kothekar E, Alavi A. Clinical Applications of Positron Emission Tomography in the Evaluation of Spine and Joint Disorders. PET Clinics: W.B. Saunders; 2019. p. 61-9.

7. Son SM, Kim K, Pak K, Kim SJ, Goh TS, Lee JS. Evaluation of the diagnostic performance of (18)F-NaF positron emission tomography/computed tomography in patients with suspected ankylosing spondylitis according to the Assessment of SpondyloArthritis International Society criteria. Spine J. 2020;20(9):1471-9.
8. de Jongh J, Verweij NJF, Yaqub M, van Denderen CJ, van der Horst-Bruinsma IE, Bot JCJ, et al. [(18)F]Fluoride PET provides distinct information on disease activity in ankylosing spondylitis as compared to MRI and conventional radiography. Eur J Nucl Med Mol Imaging. 2023;50(5):1351-9.
9. de Jongh J, Hemke R, Zwezerijnen GJC, Yaqub M, van der Horst-Bruinsma IE, van de Sande MGH, et al. (18)F-sodium fluoride PET-CT visualizes both axial and peripheral new bone formation in psoriatic arthritis patients. Eur J Nucl Med Mol Imaging. 2023;50(3):756-64.
10. Bruijnen STG, Verweij NJF, van Duivenvoorde LM, Bravenboer N, Baeten DLP, van Denderen CJ, et al. Bone formation in ankylosing spondylitis during anti-tumour necrosis factor therapy imaged by 18F-fluoride positron emission tomography. Rheumatology (Oxford). 2018;57(4):631-8.
11. Duclos V, Iep A, Gomez L, Goldfarb L, Besson FL. Pet molecular imaging: A holistic review of current practice and emerging perspectives for diagnosis, therapeutic evaluation and prognosis in clinical oncology. International Journal of Molecular Sciences: MDPI AG; 2021.
12. Arimoto MK, Nakamoto Y, Nakatani K, Ishimori T, Yamashita K, Takaori-Kondo A, et al. Increased bone marrow uptake of 18F-FDG in leukemia patients: preliminary findings. SpringerPlus. 2015;4(1).
13. Salaun PY, Gastinne T, Bodet-Milin C, Campion L, Cambefort P, Moreau A, et al. Analysis of 18F-FDG PET diffuse bone marrow uptake and splenic uptake in staging of Hodgkin's lymphoma: A reflection of disease infiltration or just inflammation? European Journal of Nuclear Medicine and Molecular Imaging. 2009;36(11):1813-21.
14. Nguyen C, Havlicek J, Duong Q, Vesely S, Gress R, Lindenberg L, et al., editors. An automatic 3D CT/PET segmentation framework for bone marrow proliferation assessment. Proceedings - International Conference on Image Processing, ICIP; 2016 2016/8//: IEEE Computer Society.
15. Meikle SR, Sossi V, Roncali E, Cherry SR, Banati R, Mankoff D, et al. Quantitative PET in the 2020s: A roadmap. Physics in Medicine and Biology. 2021;66(6).
16. Lindgren Belal S, Sadik M, Kaboteh R, Enqvist O, Ulén J, Poulsen MH, et al. Deep learning for segmentation of 49 selected bones in CT scans: First step in automated PET/CT-based 3D quantification of skeletal metastases. European Journal of Radiology. 2019;113:89-95.
17. Schott B, Weisman AJ, Perk TG, Roth AR, Liu G, Jeraj R. Comparison of automated full-body bone metastases delineation methods and their corresponding prognostic power. Physics in medicine and biology. 2023;68(3).
18. Sadik M, López-Urdaneta J, Ulén J, Enqvist O, Krupic A, Kumar R, et al. Artificial intelligence could alert for focal skeleton/bone marrow uptake in Hodgkin's lymphoma patients staged with FDG-PET/CT. Scientific Reports. 2021;11(1).
19. Zhang Z, Sejdić E. Radiological images and machine learning: Trends, perspectives, and prospects. Computers in Biology and Medicine: Elsevier Ltd; 2019. p. 354-70.
20. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50 Suppl 1:11S-20S.
21. Pan T, Mawlawi O, Luo D, Liu HH, Chi PCM, Mar MV, et al. Attenuation correction of PET cardiac data with low-dose average CT in PET/CT. Medical Physics. 2006;33(10):3931-8.
22. Townsend DW, Carney JPJ, Yap JT, Hall NC. PET/CT Today and Tomorrow. J Nucl Med. 2004;45 Supl:4S-14S.
23. Shiyam Sundar LK, Yu J, Muzik O, Kulterer O, Fueger BJ, Kifjak D, et al. Fully-automated, semantic segmentation of whole-body 18F-FDG PET/CT images based on data-centric artificial intelligence. J Nucl Med. 2022.
24. Tragardh E, Borrelli P, Kaboteh R, Gillberg T, Ulen J, Enqvist O, et al. RECOMIA-a cloud-based platform for artificial intelligence research in nuclear medicine and radiology. EJNMMI PHYSICS. 2020;7(1).
25. Wasserthal J, Meyer M, Breit H-C, Cyriac J, Yang S, Segeroth M. TotalSegmentator: robust segmentation of 104 anatomical structures in CT images. 2022.

26. Netherton TJ, Nguyen C, Cardenas CE, Chung C, Klopp AH, Colbert LE, et al. An Automated Treatment Planning Framework for Spinal Radiation Therapy and Vertebral-Level Second Check. International Journal of Radiation Oncology Biology Physics. 2022.
27. Piri R, Noddeskou-Fink A, Gerke O, Larsson M, Edenbrandt L, Enqvist O, et al. PET/CT imaging of spinal inflammation and microcalcification in patients with low back pain: A pilot study on the quantification by artificial intelligence-based segmentation. CLINICAL PHYSIOLOGY AND FUNCTIONAL IMAGING. 2022;42(4):225-32.
28. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention - MICCAI 20152015.
29. Bogduk N. Functional anatomy of the spine. Handb Clin Neurol. 2016;136:675-88.
30. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H, et al. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. Arthritis Rheum. 2006;54(8):2665-73.
31. Heuft-Dorenbosch L SA, van Tubergen A, Landewé R, van der Tempel H, Mielants H, Dougados M, van der Heijden D. Assessment of enthesitis in ankylosing spondylitis. Ann rheum Dis. 2003;62:127-32.
32. Maksymowych WP, Mallon C, Morrow S, Shojania K, Olszynski WP, Wong RL, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. Ann Rheum Dis. 2009;68(6):948-53.
33. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum. 1984;27(4):361-8.
34. Garrett S JT, Kennedy LG. et al. . A new aproach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Index. J Rheumatol. 1994;21:2286-91.
35. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2021;18(2):203-11.
36. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323-41.
37. Klein S, Staring M, Murphy K, Viergever Ma, Pluim J. elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Transactions on Medical Imaging. 2010;29(1):196-205.
38. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 2015;15:29.
39. Aydin OU, Taha AA, Hilbert A, Khalil AA, Galinovic I, Fiebach JB, et al. On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking. Eur Radiol Exp. 2021;5(1):4.
40. Kramer GM, Frings V, Hoetjes N, Hoekstra OS, Smit EF, de Langen AJ, et al. Repeatability of Quantitative Whole-Body 18F-FDG PET/CT Uptake Measures as Function of Uptake Interval and Lesion Selection in Non-Small Cell Lung Cancer Patients. J Nucl Med. 2016;57(9):1343-9.