Published in final edited form as:

Acad Radiol. 2016 October ; 23(10): 1214–1220. doi:10.1016/j.acra.2016.05.015.

Improving Spleen Volume Estimation via Computer Assisted Segmentation on Clinically Acquired CT Scans

Zhoubing Xu¹, Adam L. Gertz², Ryan P. Burke³, Neil Bansal⁴, Hakmook Kang⁵, Bennett A. Landman^{1,3,6}, and Richard G. Abramson⁶

Zhoubing Xu: zhoubing.xu@vanderbilt.edu; Adam L. Gertz: adam.l.gertz@vanderbilt.edu; Ryan P. Burke: ryan.p.burke@vanderbilt.edu; Neil Bansal: nbansal4732@gmail.com; Hakmook Kang: h.kang@vanderbilt.edu; Bennett A. Landman: bennett.landman@vanderbilt.edu; Richard G. Abramson: richard.abramson@vanderbilt.edu

¹Electrical Engineering, Vanderbilt University, Nashville, TN, 37235

²Neuroscience, Vanderbilt University, Nashville, TN, USA 37235

³Biomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

⁴Diagnostic Radiology, NYU Langone Medical Center, New York, NY, USA 10016

⁵Biostatistics, Vanderbilt University, Nashville, TN, USA 37235

⁶Radiology and Radiological Science, Vanderbilt University, Nashville, TN, USA 37235

Abstract

OBJECTIVES—Multi-atlas fusion is a promising approach for computer-assisted segmentation of anatomical structures. The purpose of this study was to evaluate the accuracy and time efficiency of multi-atlas segmentation for estimating spleen volumes on clinically-acquired CT scans.

MATERIALS AND METHODS—Under IRB approval, we obtained 294 deidentified (HIPAA-compliant) abdominal CT scans on 78 subjects from a recent clinical trial. We compared five pipelines for obtaining splenic volumes: Pipeline 1—manual segmentation of all scans, Pipeline 2—automated segmentation of all scans, Pipeline 3—automated segmentation of all scans with manual segmentation for outliers on a rudimentary visual quality check, Pipelines 4 and 5—volumes derived from a unidimensional measurement of craniocaudal spleen length and three-dimensional splenic index measurements, respectively. Using Pipeline 1 results as ground truth, the accuracy of Pipelines 2–5 (Dice similarity coefficient [DSC], Pearson correlation, R-squared, and percent and absolute deviation of volume from ground truth) were compared for point estimates of splenic volume and for change in splenic volume over time. Time cost was also compared for Pipelines 1–5.

RESULTS—Pipeline 3 was dominant in terms of both accuracy and time cost. With a Pearson correlation coefficient of 0.99, average absolute volume deviation 23.7 cm³, and 1 minute per

Correspondence author: Zhoubing Xu, Vanderbilt University EECS, 400 24th Ave So., Room 362, Nashville, TN 37235, Phone: 615-322-2338, Fax: 615-343-5459, zhoubing.xu@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

scan, Pipeline 3 yielded the best results. The second-best approach was Pipeline 5, with a Pearson correlation coefficient 0.98, absolute deviation 46.92 cm³, and 1 minute 30 seconds per scan. Manual segmentation (Pipeline 1) required 11 minutes per scan.

CONCLUSION—A computer-automated segmentation approach with manual correction of outliers generated accurate splenic volumes with reasonable time efficiency.

Keywords

Multi-atlas segmentation; spleen; computer aided diagnosis; computed tomography

Introduction

The clinical promise of computer-assisted content labeling lies in its potential to promote the extraction of quantitative morphometric information from imaging scans while minimizing time and resource requirements. The demand for such quantitative information is closely linked with the ascendancy of evidence-based medicine, which depends heavily on the statistical correlation of quantitative results from different clinical datasets [1]. In time, if computer-assisted image segmentation can enable extraction of morphometric information with sufficient accuracy and time efficiency, we may advance to the point of performing high-throughput “big data” analyses of large imaging databases. This can lead to the discovery of clinically relevant associations between imaging-based phenotypic markers and other clinical endpoints. This approach is currently applied to high-throughput genome-wide analyses in search of novel clinically-relevant genetic biomarkers [2–5].

Among a wide range of potentially extractable morphometric biomarkers from imaging scans, lesion and organ size stand out as important targets due to the historical use of lesion and organ size information as markers of disease presence, severity, and response to treatment [6–9]. Splenic volume is an intriguing biomarker to test computer-assisted segmentation techniques on because of its intersection with a wide array of disease states and the special methodologic challenges associated with segmenting this particular organ. Quantitative estimates of spleen size have been of clinical interest for decades [10], but computer-assisted labeling of the spleen has been difficult due to a wide variation between subjects regarding splenic size, shape, and geometric orientation within the abdomen [11, 12].

Due to these difficulties multi-atlas approaches have been pursued for identifying and labeling the spleen [11]. Briefly, this approach involves registering multiple previously labeled CT scans (atlases) to the target scan, selecting the reference scans from the atlases that are most similar to the target, and then using a statistical algorithm to combine the atlas labels from the reference scans to the target. This approach is promising for labeling anatomical structures in the abdomen because (a) it can be adapted to structures with variable image intensity and appearance, (b) it is robust to different scanning parameter settings while also usually free from interactive initialization (e.g., manual positioning of region of interest), and (c) it can be augmented by transferring newly segmented scans into existing atlases, thus allowing the algorithm to become “smarter” over time [13].

The purpose of this study was to evaluate the accuracy and time efficiency of the multi-atlas segmentation technique for estimating spleen volumes on clinically-acquired CT scans. Since preliminary work suggested that multi-atlas labeling performs exceptionally well in a majority of scans but poorly in a small subset of outliers, we investigated two “pipelines” for using multi-atlas labeling to generate splenic volumes. One pipeline consisted of computer-assisted multi-atlas segmentation on all scans, while the other pipeline consisted of computer-assisted segmentation supplemented by manual segmentation for scans that failed a rudimentary and high-level visual quality check (Figure 1). We also evaluated two more traditional methods for quickly estimating splenic volume: a single-dimension craniocaudal measurement and a three-dimensional splenic index. These were both transformed into volume estimates by linear regression equations found in the literature [14]. We compared the accuracy of these four approaches for estimating splenic volume against ground truth values established by manual segmentation. Furthermore, since intra-subject changes in splenic volume over time may be clinically relevant biomarkers in certain settings [15, 16], we also compared the accuracy of these four approaches for estimating the change in splenic volume from prior measures for the same subject. Finally, we compared the average time it took for all pipelines to generate splenic volumes.

Methods and Materials

Data acquisition

Under institutional review board supervision and in compliance with HIPAA requirements, we obtained all contrast-enhanced CT scans performed at our institution for a recently completed phase 2 metastatic colorectal cancer clinical trial. The clinical trial cohort included 78 unique patients (Table 1), all of whom had from one to eight scans during the trial for a total of 294 abdominal CT scans. Imaging was acquired on different scanners, with varying fields of view ($280 \times 280 \times 175$ to $500 \times 550 \times 780$ mm³), in-plane resolution (0.54×0.54 to 0.98×0.98 mm²), slice thickness (1.5 mm to 7.0 mm), and in portal venous phase with IV contrast amounts (70 to 120 mL iohexol 350, GE Healthcare, Waukesha, WI). Using the DCM2NII tool from the MRICron package [17], we converted the scans from DICOM into NIFTI format for image processing. Seventy-five of the scans have been reported in previous technical studies [12, 18, 19]. We performed our first clinical study with a larger dataset in this manuscript.

Manual segmentation (Pipeline 1)

An experienced research associate manually segmented the spleen on all 294 scans to obtain baseline splenic volumes. Briefly, each scan was opened in the MIPAV software (NIH, Bethesda, MD [20]), and spleen outlines were delineated on every axial slice using an electronic pen and tablet device (Wacom Intuos, Vancouver, WA). Spleen volumes were derived by filling the regions enclosed by the spleen outlines and multiplying the counts of the labeled spleen voxels by the voxel size of the scan. A radiologist verified all splenic contours on a volumetric basis. The time from starting the labeling process to calculating the splenic volume was tracked for each scan.

Additionally, a subset of 74 scans was manually labeled by a second research associate in order to assess the inter-rater reliability of the manual segmentation process. This second research associate followed the same outline-filling and volume-measuring process as described above.

Computer-assisted segmentation (Pipeline 2)

Automated segmentation of the spleen was performed for all scans using a recently published multi-atlas label fusion technique [18]. Briefly, 65 scans were randomly selected from the baseline scans (i.e., first time point of the clinical trial) to serve as the reference atlases. For each target image of the 294 CT scans, 65 or 64 non-rigid registrations [21] were performed through which the manually labeled spleens were propagated from the atlases to the target. This excluded the target itself if it was within the atlases. Atlas selection and statistical label fusion were performed to yield the estimated spleen segmentation by combining the registered spleen labels. Specifically, the atlases were iteratively evaluated and selected based on the volumetric overlap they had with the intermediate result across the remaining atlases using an Expectation-Maximization scheme. The intermediate result was a majority-voted fusion of the remaining registered atlases adjusted by the spleen likelihood on the target scan, where the spleen likelihood was inferred from a Gaussian mixture model (pre-trained with 10 of the 65 atlases) of the image contexts [12]. Joint label fusion [22] was employed on the selected atlases to yield the final segmentation. The voxel-wise decisions were based on a weighted voting of the atlases where the weights were derived on each voxel for each atlas to minimize the correlated errors across the atlas images in local neighborhoods. The resultant labeled regions (Figure 2) were then used to calculate the estimated splenic volume by multiplying the counts of the labeled spleen voxels by the voxel size of the scan. The total time of manual interaction for running the spleen segmentation and calculating the splenic volume for each scan was tracked.

Computer-assisted segmentation with manual labeling for outliers (Pipeline 3)

Pipeline 3 simulated a workflow in which manual segmentation was used only for those computer-segmented scans that failed a rudimentary visual quality check. Performed on all computer-segmented scans, the quality check consisted of overlaying the spleen segmentation result on a single axial, coronal, and sagittal plane that sliced through the middle of the spleen (Figure 3). Based on qualitative judgment, a scan failed the quality check if its segmentation on at least one slice appeared to capture less than 50% of the true underlying structure. If a scan passed the quality check, the computer-assisted splenic volume was accepted.

If a scan failed the quality check, the splenic volume from manual segmentation (Pipeline 1) was used. Time per scan for this pipeline was computed as the average time required to perform the quality check plus the average time it took to segment the images and generate a splenic volume (including all required manual interactions).

Unidimensional and splenic index measurements (Pipelines 4 and 5)

An experienced research associate extracted linear measurements of spleen size on all scans while following the earlier work of Bezerra et al. [14], who compared different linear spleen

measurements and indices and presented transfer equations correlating these measurements with splenic volumes. In Pipeline 4, a unidimensional measurement of craniocaudal spleen length (L) was extracted as described in [14] by multiplying the number of axial sections where the spleen was visualized with the thickness of the sections (Figure 2). Spleen length was then converted into a splenic volume using the linear regression equation $L = (0.0126 \times \text{Volume}) + 5.8006$ [14]. In Pipeline 5, a three-dimensional splenic index (SI) was extracted as described in [14] by multiplying craniocaudal spleen length (L) by the maximum axial plane width (W) on any section (typically in oblique sagittal plane, see Figure 2) by the maximum axial plane thickness (T) on any section (perpendicular to W, see Figure 2). The resultant splenic index ($SI = L \times W \times T$) was then converted into a splenic volume using the linear regression equation $SI = (2.7036 \times \text{Volume}) + 0.4776$ [14]. The times to generate splenic volumes were tracked separately for Pipelines 4 and 5.

Statistical analysis

The following set of metrics were used to evaluate segmentation accuracy: Dice similarity coefficient (DSC) [23], Pearson correlation, R-squared, and absolute and percent deviation of splenic volumes from ground truth. First, these accuracy metrics were computed for the subset of 74 scans in Pipeline 1 that underwent segmentation by two independent research associates in order to assess the inter-rater reliability of the manual segmentation process in Pipeline 1. Then, the accuracy measures were computed for each Pipeline using the complete set of 274 manually segmented scans (Pipeline 1) as the ground truth to evaluate the ability of Pipelines 2–5 to generate point estimates of splenic volume. The accuracy metrics were again computed for Pipelines 2–5 using the information from Pipeline 1 as the ground truth to evaluate the ability of Pipelines 2–5 to estimate changes in splenic volume between successive measures for an individual subject. These computations were performed both for change in splenic volume from baseline (i.e., first time point of the clinical trial) and for change from the most recent prior scan.

One-tail Wilcoxon signed rank tests were performed to compare the accuracy measures (DSC and absolute deviation) if applicable. The significance tests were performed again for the time cost between Pipelines 3–5. Bland-Altman plots were constructed for each of the Pipelines 2–5 to depict the agreement against Pipeline 1 on the point estimates of splenic volume (Figure 4).

IRB approval

This study was approved by the Vanderbilt University Institutional Review Board on August 15, 2012.

Results

Inter-rater reliability of manual segmentation

Metrics for Pipeline 1 inter-rater reliability are presented in Table 2. The manual spleen segmentation demonstrated high reproducibility with an average DSC of 0.96 between labels from two individual raters. The extracted splenic volume sizes were also highly correlated in terms of the Pearson correlation and R-squared.

Point estimates of splenic volume

For Pipeline 2 the computer-assisted segmentation achieved a DSC of 0.90 ± 0.11 against the manual segmentation. Twenty out of 294 scans were below 0.70 in DSC. Twenty-two out of 294 scans failed the visual quality check and required re-labeling when performing Pipeline 3. This approximately matched the number of scans with low DSC values. With outlier correction, Pipeline 3 achieved a DSC of 0.93 ± 0.07 and a substantially better agreement in spleen volume measurement than Pipeline 2 in terms of Pearson correlation (0.98 vs. 0.72) and R-Squared (0.98 vs. 0.51).

Splenic volume estimation using 3-D splenic index (Pipeline 5) performed better than 1-D splenic length (Pipeline 4). Close values of Pearson correlation and R-squared were observed between Pipelines 3 and 5, although Pipeline 3 yielded a significantly smaller ($p < 0.05$) absolute deviation of estimated volume from ground truth. The Bland-Altman plots (Figure 4) show that Pipeline 3 provided the best agreement with manual segmentation among Pipelines 2–5.

Estimates of change in splenic volume

For both the change from baseline and most recent scan, Pipeline 3 produced the most accurate results and Pipeline 5 produced the second most accurate results. In general, the estimates for change in splenic volume were not as accurate as the single point estimates for splenic volume.

Time costs

Table 2 reports the manual interaction time for each method. Pipeline 3 required 50 seconds for manual labeling, 5 seconds for running the segmentation program, and 5 seconds for quality assurance. For Pipelines 4 and 5, the average time to search for the scan (by medical record number and scan date), load the scan, cache the images for analysis, and close the scan was 39 seconds.

Discussion

Main contributions

This study demonstrates the feasibility of using computer-automated segmentation to extract spleen volumes on large CT datasets. Pipeline 3 – i.e., computer segmentation for a majority of scans and manual segmentation for a minority of easily recognized outliers – was dominant among Pipelines 2–5 in terms of both accuracy and time cost, and achieved a Pearson correlation coefficient of 0.99 and an average absolute deviation of 23.70 cm^3 from ground truth spleen volumes at a time cost of only 1 minute per scan. This approach was superior to computer segmentation (Pearson correlation 0.72, absolute deviation 64.66 cm^3 , time cost 5 seconds per scan), unidimensional measurement (Pearson correlation 0.86, absolute deviation 111.02 cm^3 , time cost 1 minute 5 seconds per scan), and 3-D splenic index measurement (Pearson correlation 0.98, absolute deviation 46.92 cm^3 , time cost 1 minute 30 seconds per scan). These results show that a hybrid approach combining fully automated processing with manual correction can enable generation of accurate spleen volumes with reasonable time efficiency.

In reviewing the computer-automated segmentations, we found that 90% of the spleen volumes from Pipeline 2 were highly accurate, but about 10% of outliers were highly inaccurate. Typical outliers include spleens with large volumes (at least twice the average), and those about other organs (liver and left kidney). These outlier segmentations were easily spotted with a rudimentary visual quality check. When these outliers were manually segmented the average time for manual interaction was still less than that of Pipelines 4 and 5.

Potential improvement

The presented computer-assisted segmentation approach achieved spleen volume accuracies comparable to other recent related works. Wolz et al. used a hierarchical scheme of multi-atlas segmentation and achieved a DSC of 0.92 for spleen [24]. Linguraru et al. presented a spleen segmentation tool with an average DSC of 0.95 using probabilistic atlases followed by the correction of shape, location, and contrast enhancement [11]. Okada et al. modeled the inter-organ relationship of shape and location in the abdomen and yielded the spleen segmentation with a DSC of 0.93 [25]. We note that unlike other prior attempts, our dataset was comprised of clinically acquired CT scans at conventional doses.

The automatic segmentation is the crux of the efficiency and robustness for the splenic volume estimation in this study. The presented multi-atlas label fusion approach can be further improved from two perspectives. First, pre-localizing the spleen region of interest (ROI) can largely decrease the registration failures and thus enhance the segmentation performances. This enhanced version has been implemented where better accuracies were observed [19]. Second, a prospective feedback loop can be constructed where datasets failing the quality check would be appended to the atlas set. For the multi-atlas label fusion technique, it is critical to select atlases similar to the scan, especially when there are large shape and context variations in the structure of interest. A growing atlas set can benefit spleen segmentation on later datasets by providing larger coverage of spleen variations to capture the extreme cases. For example, in this study most of the scans requiring re-labeling had a large spleen volume, which was beyond the capability of the current atlas set. It can be expected that an augmented atlas set including these scans could capture large spleens better in the future.

Implication for patient care

We see a variety of potential clinical applications for this technique. Spleen volumes are relevant biomarkers for the presence of many disease processes (e.g., small spleen in sickle cell disease, large spleen in portal hypertension, certain hematologic malignancies, infiltrative processes such as Gaucher's and glycogen storage diseases, and infections such as HIV and infectious mononucleosis). Tracking spleen volume changes over time has a potential clinical role for monitoring disease progression as a prognostic indicator for disease complications (e.g., esophageal varices), and as a biomarker for treatment response (e.g., response to therapy in myelofibrosis [16, 26]). While spleen volumes are currently most often described on clinical CT reports in qualitative terms, quantification offers the potential for increased diagnostic accuracy, decreased subjectivity of interpretation, and higher sensitivity for subtle and potentially clinically relevant changes. Computer-assisted

segmentation applied to large population datasets has the potential to improve on currently accepted reference values for abnormal spleen volumes adjusted for covariates including age, gender, and body mass index. Finally, our proposed multi-atlas fusion approach provides information not just on splenic volume, but also on geometric features associated with the shape of the organ itself. This can provide for interesting and relevant correlations when applied to high-throughput “big data” phenotype-clinical analyses.

Summary

In summary, we presented a hybrid computer-assisted segmentation method with manual segmentation of outliers that provides accurate estimation of spleen volumes at a reasonable time cost. Our method opens the opportunity for a variety of potential clinical applications. We expect our multi-atlas fusion algorithm to improve with increasing size of the reference atlas set which will decrease the need for manual segmentation on outliers. We envision a future in which this technique may be incorporated into PACS software for routine reporting of quantitative morphometric information, not only for spleen but for other organs and for discrete lesions as well. The path for clinical translation likely involves incorporation first into clinical trials and specialized core laboratories, and subsequently into broad clinical use as time efficiencies improve.

Acknowledgments

The authors thank Naresh Nandakumar for the English language edit. This research was supported by NIH 1R03EB012461, NIH 2R01EB006136, NIH R01EB006193, ViSE/VICTR VR3029, NIH UL1 RR024975-01, NIH UL1 TR000445-06, NIH P30 CA068485, and AUR GE Radiology Research Academic Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

References

1. Sullivan DC. Imaging as a quantitative science. *Radiology*. 2008; 248(2):328–32. [PubMed: 18641239]
2. Parkinson H, et al. ArrayExpress update--an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res*. 2011; 39(Database issue):D1002–4. [PubMed: 21071405]
3. Barrett T, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res*. 2009; 37(Database issue):D885–90. [PubMed: 18940857]
4. Bourgon R, et al. High-throughput detection of clinically relevant mutations in archived tumor samples by multiplexed PCR and next-generation sequencing. *Clin Cancer Res*. 2014; 20(8):2080–91. [PubMed: 24573554]
5. Roychowdhury S, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med*. 2011; 3(111):111ra121.
6. Wolf AD, Lavine JE. Hepatomegaly in neonates and children. *Pediatr Rev*. 2000; 21(9):303–10. [PubMed: 10970452]
7. O'Reilly RA. Splenomegaly in 2,505 patients at a large university medical center from 1913 to 1995. 1963 to 1995: 449 patients. *West J Med*. 1998; 169(2):88–97. [PubMed: 9735689]
8. Felker GM, et al. Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. *N Engl J Med*. 2000; 342(15):1077–84. [PubMed: 10760308]
9. Eisenhauer EA, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009; 45(2):228–47. [PubMed: 19097774]

10. Dittrich M, et al. Sonographic biometry of liver and spleen size in childhood. *Pediatr Radiol*. 1983; 13(4):206–11. [PubMed: 6888991]
11. Linguraru MG, et al. Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. *Medical physics*. 2010; 37(2):771–783. [PubMed: 20229887]
12. Xu, Z.; Asman, AJ.; Shanahan, PL., et al. Medical image computing and computer-assisted intervention–MICCAI 2014. Springer; 2014. SIMPLE is a good idea (and better with context learning); p. 364–371.
13. Wolz R, et al. LEAP: learning embeddings for atlas propagation. *Neuroimage*. 2010; 49(2):1316–1325. [PubMed: 19815080]
14. Bezerra AS, et al. Determination of splenomegaly by CT: is there a place for a single measurement? *American Journal of Roentgenology*. 2005; 184(5):1510–1513. [PubMed: 15855107]
15. Overman MJ, et al. Oxaliplatin-mediated increase in spleen size as a biomarker for the development of hepatic sinusoidal injury. *J Clin Oncol*. 2010; 28(15):2549–55. [PubMed: 20406923]
16. Tefferi A, et al. Revised response criteria for myelofibrosis: International Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT) and European LeukemiaNet (ELN) consensus report. *Blood*. 2013; 122(8):1395–8. [PubMed: 23838352]
17. Rorden C, Brett M. Stereotaxic display of brain lesions. *Behavioural neurology*. 2000; 12(4):191–200. [PubMed: 11568431]
18. Xu Z, Burke RP, Lee CP, et al. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Med Image Anal*. 2015; 24:18–27. [PubMed: 26046403]
19. Xu Z, Gertz AL, Baucom RB, et al. Improving Multi-Atlas Abdominal Organ Segmentation via Organ-Wise Normalization. *MICCAI 2015 Workshop and Challenge: Multi-Atlas Labeling Beyond Cranial Vault*. 2015
20. McAuliffe, MJ., et al. Medical image processing, analysis and visualization in clinical research. *Proceedings of the 14th IEEE Symposium on Computer-Based Medical Systems*; IEEE; 2001.
21. Rueckert D, et al. Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging IEEE Transactions on*. 1999; 18(8):712–721.
22. Wang H, et al. Multi-atlas segmentation with joint label fusion. *Pattern Analysis and Machine Intelligence IEEE Transactions on*. 2013; 35(3):611–623.
23. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945; 26(3):297–302.
24. Wolz R, et al. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *Medical Imaging IEEE Transactions on*. 2013; 32(9):1723–1730.
25. Okada T, et al. Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. *Medical image analysis*. 2015; 26(1):1–18. [PubMed: 26277022]
26. Harrison C, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N Engl J Med*. 2012; 366(9):787–98. [PubMed: 22375970]

Figure 1.
Overview of pipelines for estimating spleen volumes.

Figure 2.

Illustration of the required measurements from different pipelines for estimating the spleen volume. Pipeline 2 and 3 extract the whole spleen volume while pipelines 4 and 5 measure splenic diameters along different axes.

Figure 3.

Quality assurance of the computer assisted segmentation in Pipeline 3 was performed by overlaying the spleen segmentation result on single axial, coronal, and sagittal CT slices through the middle of the spleen. Upper row: a successful case where the automated labels were used. Lower row: a failure case where manual correction was required.

Figure 4.

Bland-Altman plots for different spleen volume estimation methods using pipelines 2–5. On each plot, the horizontal axis represents the mean volume between the ground truth and the estimation. The vertical axis indicates the difference in volume from the ground truth to the estimation. The mean in difference, and a confidence interval of ± 1.96 standard deviation (SD) are highlighted.

Table 1

Clinical parameters for the tested 78 subjects

Parameter	Range	N (%)
Age	<40	4 (5)
	40–49	14 (18)
	50–59	23 (29)
	60–69	23 (29)
	70–79	10 (13)
	80+	4 (5)
	Total	78 (100)
Gender	Male	45 (58)
	Female	33 (42)
	Total	78 (100)
BMI	<20	3 (4)
	20–24.9	24 (31)
	25–29.9	20 (26)
	30–34.9	19 (24)
	35+	11 (14)
	Not available	1 (1)
	Total	78 (100)

Table 2

Accuracy and Time Results for Each Pipeline

Result	Manual segmentation (Pipeline 1) ^a	Computer segmentation for all scans (Pipeline 2)	Computer segmentation with manual segmentation of outliers (Pipeline 3)	1-D length (Pipeline 4)	3-D splenic index (Pipeline 5)
Accuracy (point estimates)					
Dice similarity coefficient	0.96 ± 0.01	0.90 ± 0.11	0.93 ± 0.07 ^{**}	N/A	N/A
Pearson correlation	0.9997	0.7151	0.9888	0.8613	0.9765
R-squared	0.9993	0.5114	0.9778	0.7435	0.9535
Absolute deviation of volume from ground truth (cm ³)	7.25 ± 7.29	64.66 ± 263.93	23.70 ± 48.98 ^{**}	111.02 ± 168.48	46.92 ± 66.37
Accuracy (change from baseline)					
Pearson coefficient	N/A	0.5556	0.8741	0.4839	0.8178
R-squared	N/A	0.3087	0.7641	0.2717	0.7437
Absolute deviation of volume change from ground truth (cm ³)	N/A	46.70 ± 84.86	28.24 ± 49.55 ^{**}	81.62 ± 107.82	38.11 ± 68.01
Accuracy (change from most recent prior)					
Pearson coefficient	N/A	0.6532	0.8094	0.4825	0.7694
R-squared	N/A	0.4267	0.6551	0.2590	0.6597
Absolute deviation of volume change from ground truth (cm ³)	N/A	38.04 ± 67.07	26.33 ± 52.12 ^{**}	64.28 ± 81.64	33.95 ± 59.34
Time cost					
Manual interaction time (averaged per scan)	11 min	5 sec	1 min [*]	1 min 5 sec	1 min 30 sec

^a measured between the labels from two independent raters on a subset of 74 scans

^{*} Dominant at a p<0.05 using one-tail Wilcoxon signed rank test between Pipelines 3–5

^{**} Dominant at a p<0.05 using one-tail Wilcoxon signed rank test ((comparing to the second best)