




# Development of a volumetric pancreas segmentation CT dataset for AI applications through trained technologists: a study during the COVID 19 containment phase

Garima Suman<sup>1</sup> · Ananya Panda<sup>1</sup> · Panagiotis Korfiatis<sup>1</sup> · Marie E. Edwards<sup>1</sup> · Sushil Garg<sup>2</sup> · Daniel J. Blezek<sup>1</sup> · Suresh T. Chari<sup>3</sup> · Ajit H. Goenka<sup>1</sup> 

Received: 24 July 2020 / Revised: 26 August 2020 / Accepted: 3 September 2020 / Published online: 16 September 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

**Purpose** To evaluate the performance of trained technologists vis-à-vis radiologists for volumetric pancreas segmentation and to assess the impact of supplementary training on their performance.

**Methods** In this IRB-approved study, 22 technologists were trained in pancreas segmentation on portal venous phase CT through radiologist-led interactive videoconferencing sessions based on an image-rich curriculum. Technologists segmented pancreas in 188 CTs using freehand tools on custom image-viewing software. Subsequent supplementary training included multimedia videos focused on common errors, which were followed by second batch of 159 segmentations. Two radiologists reviewed all cases and corrected inaccurate segmentations. Technologists' segmentations were compared against radiologists' segmentations using Dice-Sorenson coefficient (DSC), Jaccard coefficient (JC), and Bland–Altman analysis.

**Results** Corrections were made in 71 (38%) cases from first batch [26 (37%) oversegmentations and 45 (63%) undersegmentations] and in 77 (48%) cases from second batch [12 (16%) oversegmentations and 65 (84%) undersegmentations]. DSC, JC, false positive (FP), and false negative (FN) [mean (SD)] in first versus second batches were 0.63 (0.15) versus 0.63 (0.16), 0.48 (0.15) versus 0.48 (0.15), 0.29 (0.21) versus 0.21 (0.10), and 0.36 (0.20) versus 0.43 (0.19), respectively. Differences were not significant ( $p > 0.05$ ). However, range of mean pancreatic volume difference reduced in the second batch [– 2.74 cc (min – 92.96 cc, max 87.47 cc) versus – 23.57 cc (min – 77.32, max 30.19)].

**Conclusion** Trained technologists could perform volumetric pancreas segmentation with reasonable accuracy despite its complexity. Supplementary training further reduced range of volume difference in segmentations. Investment into training technologists could augment and accelerate development of body imaging datasets for AI applications.

**Keywords** Deep learning · Data curation · Artificial intelligence · COVID-19

✉ Ajit H. Goenka  
goenka.ajit@mayo.edu

Garima Suman  
suman.garima@mayo.edu

Ananya Panda  
panda.ananya@mayo.edu

Panagiotis Korfiatis  
korfiatis.panagiotis@mayo.edu

Marie E. Edwards  
edwards.marie@mayo.edu

Sushil Garg  
garg.sushil@mayo.edu

Daniel J. Blezek  
blezek.daniel@mayo.edu

Suresh T. Chari  
STChari@mdanderson.org

<sup>1</sup> Department of Radiology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

<sup>2</sup> Department of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

<sup>3</sup> Department of Gastroenterology, Hepatology and Nutrition, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA

## Introduction

The development of artificial intelligence (AI) models through supervised training of deep learning algorithms for medical imaging applications requires large training datasets with high-quality labels [1, 2]. Such datasets have been typically obtained through manual annotations of images by expert radiologists or trained image analysts. However, the process of manual segmentation and labeling on cross-sectional imaging is labor-intensive and not scalable. Recruitment of experts for segmentation and labeling also makes the generation of such datasets an expensive investment. Therefore, there is a need for alternate approaches to circumvent this bottleneck to accelerate the generation of labeled datasets for training and eventual clinical deployment of reliable AI models.

Technologists are one of the key stakeholders in the medical imaging workflow [3]. They often gain working knowledge of cross-sectional anatomy and skillsets for image processing as part of their training and clinical assignments. In fact, some technologists also go on to become members of imaging core labs at many institutions. These attributes suggest that technologists can be a potential group to generate labeled body imaging datasets for AI applications. An advantage of training technologists in image annotation tasks is that the data do not have to leave institutional security firewalls. Secondly, these trained technologists could be integrated into the data annotation pipelines of multiple other body imaging AI projects. However, to the best of our knowledge, the feasibility of this approach has not been evaluated.

Our group is developing AI-powered workflow modules to address the unmet needs in patients with pancreatic diseases. The pancreas is a solid retroperitoneal organ that can be hard to segment because of its small size, complex anatomy, and variability in location, morphology, and attenuation [4]. Furthermore, the variable degrees of peripancreatic fat, contrast enhancement, and subadjacent iso-attenuating structures such as collapsed bowel can further confound delineation of its exact boundaries [5–7]. These factors make manual segmentation of the pancreas a challenge and at least partly contribute to the underutilization of pancreas morphometrics and radiomics in both endocrine and exocrine diseases despite promising results [8–11]. Therefore, there is a need for large volume segmented datasets to develop and test production-scale AI models for automated pancreas segmentation.

During the coronavirus disease of 2019 (COVID-19) containment phase, similar to other institutions [12, 13], we faced a situation of reduced clinical imaging volumes and redundancy of staff such as technologists due to voluntary deferral of all elective clinical care by our institution.

We decided to leverage this opportunity to assess whether the skillsets of technologists could be augmented through focused training to create a CT dataset of segmented normal pancreas for AI applications in body imaging. The purpose of this project was to evaluate the performance of technologists vis-à-vis radiologists for volumetric pancreas segmentation after initial training and to assess the impact of focused supplementary training on their performance.

## Methods

### Patient cohort

The project was conducted as a part of an Institutional Review Board (IRB)-approved and Health Information Portability and Accountability Act-compliant study. The requirement for informed patient consent had been waived by the IRB due to the retrospective study design. We randomly selected 347 contrast-enhanced CT scans on the basis of a statement of a negative or unremarkable pancreas in the original radiologist's report. This was subsequently verified during manual pancreas segmentation by two radiologists (AP and GS with 7 and 3-years of post-residency experience, respectively). For each CT study, an axial portal venous phase series ( $\leq 3$ -mm slice thickness) was identified and confirmed with the use of information from the series name and DICOM header. All CT studies were de-identified by anonymization of Digital Imaging and Communication in Medicine (DICOM) tags utilizing Clinical Trial Processor [14]. These anonymized CT datasets were extracted and converted into the Neuroimaging Informatics Technology Initiative (NIFTI) format. These anonymized datasets were stored in an offline shared folder for radiologists' review on a free and open-source software package for image analysis and scientific visualization [3D Slicer<sup>®</sup> (version 4.11.0)] [15].

### Technologists' training and segmentation

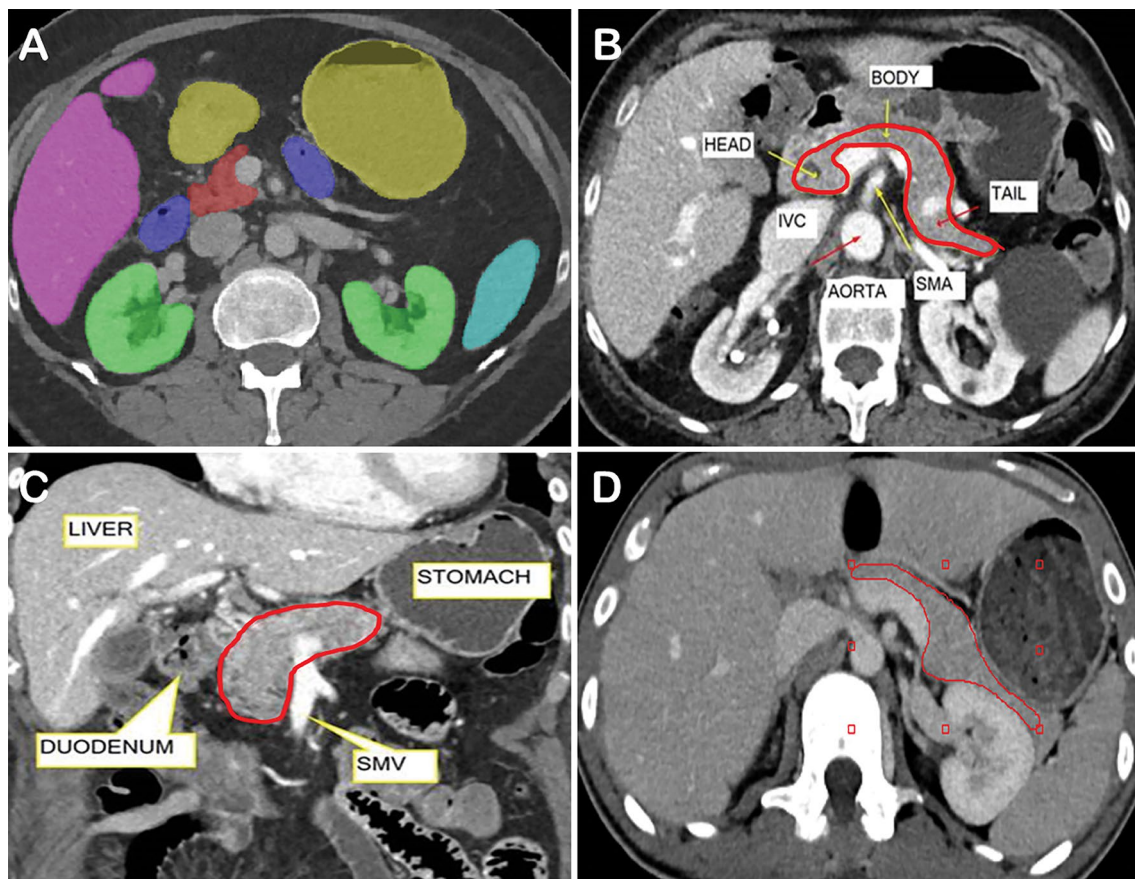
Between March and April 2020, 22 CT and MRI technologists volunteered to participate in this project. These technologists were not familiar with the 3D Slicer<sup>®</sup> software that was being used by radiologists. Therefore, we decided to train the technologists for pancreas segmentation on our enterprise custom image-viewing software (QREADS). This custom enterprise software is routinely used by the technologists to review images as part of their regular clinical work. However, they were not familiar with the image annotation tools that this software provides. To address this, a standard operating procedure (SOP) document and a 20-min training video that demonstrated steps for image display and review with the use of standard viewing tools (zoom, contrast, scroll, pan, etc.) and image annotation on a slice-by-slice

basis using freehand annotation tools were created. The SOP document contained details of the various steps involved such as image retrieval, data organization, links to access the training material, case assignment, data reporting, and quality control. To augment knowledge of the technologists, a curriculum document with infographics focused on pancreas segmentation (Fig. 1) was created by the radiologists over a period of 2 days. The topics covered in the curriculum document included an overview of project goals and an image-rich multiplanar depiction of pancreatic anatomy on CT, common anatomic variations (e.g., variations in the location of the pancreas, lipomatosis, variable pancreatic parenchymal enhancement), and relevant CT artifacts (e.g., partial volume effect, motion artifacts, streak artifacts from embolization coils).

These training documents were reviewed with the technologists in four radiologists-led interactive virtual instructional sessions of 1-hour duration each. All these instructional sessions were recorded. A recording of the session along with screencast of the workflow and training module

documents were shared with the technologists through a shared folder on the institutional intranet. Each participant technologist was required to document completion of the required training by signing off an online verification form. Finally, the technologists were also given institutional access to an interactive e-anatomy atlas ([www.imaios.com](http://www.imaios.com)) for additional but optional self-directed learning.

Following this training, an initial batch of 188 CT studies was randomly selected from the master dataset of 347 studies and was retrieved on the enterprise software. The technologists performed volumetric pancreas segmentation on a slice-by-slice basis using freehand segmentation tools over a period of 14 workdays. Queries of the technologists during this initial segmentation process were answered by radiologists through emails. These segmentations were saved, exported offline, and converted to NIFTI format. Two radiologists (AP and GS) subsequently reviewed the volumetric CT datasets and the technologists' segmentations on 3D Slicer®. These two radiologists repeated those pancreatic segmentations that were either an undersegmentation (any



**Fig. 1** Images from training material: Color-coded depiction of abdominal organs (a) on an axial CT image (pancreas: red; liver: purple; kidneys: light green; stomach: yellow; small bowel: blue, and spleen: cyan). Depiction of pancreas outline in red with labeled

subadjacent anatomical structures on axial (b) and coronal (c) CT images. Tracing of pancreas outline on enterprise custom image-viewing software using freehand tools (d). The smaller red squares are artifactually generated by the software with any outline task

part of pancreatic parenchyma left out) or an oversegmentation (any part of subadjacent anatomy included) error. Repeat segmentation was done by the radiologists with the use of boundary-points based segmentation mode of the AI-assisted segmentation module (NVIDIA) in 3D Slicer®. As part of this mode, radiologists placed input points at the perimeter of the pancreas on multiple planes (i.e., axial, coronal, and sagittal). The AI-assisted segmentation was then manually fine-tuned by the radiologists. Based on the radiologists' repeat segmentations as the ground truth, the technologists' segmentation errors were also quantified on a pixel-wise basis as either false positives (FP), i.e., the percentage of pixels segmented by technologists but not by radiologists, a measure of oversegmentation, and false negatives (FN), i.e., the percentage of pixels not included by technologists but present in radiologists' segmentations, a measure of undersegmentation (Fig. 2). Radiologists also subjectively noted the most common causes of segmentation errors.

Based on the assessment of segmentations performed in this batch, supplementary training material was created to highlight on common segmentation errors. This material included videos of representative samples of radiologists' corrected segmentations overlaid on the technologists' original segmentations. These segmentations were differently color-coded to highlight the pancreatic region(s) that were commonly being left out or the extra-pancreatic anatomy that was often being included by the technologists (Fig. 2). Additional presentations depicting the subjacent anatomy using different color codes were also prepared to improve understanding of locoregional anatomy. These supplementary materials were reviewed through virtual video meetings and were also made available to the technologists through the common shared folder. Subsequent to this supplementary training, the technologists segmented the pancreas in the second batch of another 159 CT studies over a period of 9 workdays. Additional queries were addressed via emails. Finally, the second batch of segmentations was reviewed and evaluated by the radiologists similar to the first batch.

Both batches of segmentations were performed by the technologists in the downtime during regular clinical duties. There was no provision of additional remuneration for participation in this project.

## Statistical analyses

Statistical analyses were performed with Python software (version 3.7.8; Python Software Foundation, Wilmington, Del) by using the Scikit-learn library (version 0.23.1) [16]. For the segmentations performed by technologists that were deemed inaccurate, the segmentations repeated by the radiologists were the ground truth. The original technologists' and revised radiologists' segmentations were compared using similarity metrics such as Dice–Sorenson coefficient (DSC) and Jaccard coefficient (JC). Semantic uncertainty was assessed by FP and FN rates. To evaluate the impact of supplementary training, the proportion of cases that needed no revision, oversegmentation and undersegmentation errors were compared between the two batches of segmentations using the Chi-square test for proportions. The DSC, JC, FP, and FN before and after supplementary training were compared using Kruskal–Wallis tests. Bland–Altman analysis was performed to evaluate the mean pancreatic volume difference (technologists' segmentation minus ground truth segmentation) versus the means of pancreatic volumes before and after supplementary training [17]. A  $p$  value  $< 0.05$  was considered statistically significant.

## Results

Of the initial batch of 188 segmentations, 117 (62%) were deemed accurate by radiologists and 71 (38%) had to be repeated due to segmentation errors. Undersegmentation accounted for the majority of the errors, 45/71 (63%), while the remainder [26/71 (37%)] were oversegmentation errors. Subjectively, the undersegmentation errors were commonly



**Fig. 2** Evaluation of technologists' segmentation: Color-coded areas represent correct segmentation (blue), incorrect segmentation (red), and overlap between technologists' and radiologists' segmentation (purple). Example of accurate segmentation (a); exclusion of a por-

tion of pancreatic head resulted in undersegmentation error or false negative (b), and inclusion of duodenum within the segmentation resulted in an oversegmentation error or false positive (c)

due to missing terminal portions of the head or tail of the pancreas and not including additional lobulations of pancreatic tissue separate from main pancreatic parenchyma. Oversegmentation errors were commonly due to the inclusion of iso-attenuating adjacent duodenum, collapsed jejunum, or the stomach. The DSC was  $0.63 \pm 0.15$  and JC was  $0.48 \pm 0.15$  (mean  $\pm$  SD). The FP rate was  $0.29 \pm 0.21$  and FN rate was  $0.36 \pm 0.20$  (mean  $\pm$  SD) (Table 1). From Bland–Altman analysis (Fig. 3a), mean pancreatic volume difference (technologists' segmentation minus ground truth segmentation) was  $-2.74$  cc (minimum  $-92.96$  cc, maximum  $87.47$  cc).

Out of the 159 segmentations performed in the second batch after supplementary training, 82 (52%) were deemed accurate and 77 (48%) segmentations had to be repeated. Oversegmentations were seen in 12/77 (16%) cases while 65/77 (84%) were undersegmentations. The causes of oversegmentations and undersegmentations were similar to those in the first batch. The DSC was  $0.63 \pm 0.16$  and JC was  $0.48 \pm 0.15$  (mean  $\pm$  SD). The FP rate was  $0.21 \pm 0.10$

and FN rate was  $0.43 \pm 0.19$ , (mean  $\pm$  SD) (Table 1). From Bland–Altman analysis (Fig. 3b), mean pancreatic volume difference (technologists' segmentation minus ground truth segmentation) was  $-23.57$  cc (minimum  $-77.32$  cc, maximum  $30.19$  cc).

There was no difference in the proportion of accurate segmentations between the first and the second batch of technologists' segmentations (62% in the first batch and 52% in the second batch,  $p=0.06$ ). The trend of decline in the proportion of accurate segmentations in the second batch was primarily due to a relative increase in the share of undersegmentation errors (63% in the first batch and 84% in the second batch,  $p=0.003$ ). Conversely, there was a decrease in the share of oversegmentation errors (37% in the first batch and 16% in the second batch,  $p=0.003$ ). However, the range of mean pancreatic volume difference after supplementary training was lower than in the first batch ( $-77.32$  to  $30.19$  cc compared to  $-92.96$  to  $87.47$  cc in the first batch). There was no difference in DSC ( $p=0.61$ ), JC ( $p=0.61$ ), FP ( $p=0.07$ ), and FN rates ( $p=0.12$ ) between the two batches (Fig. 4).

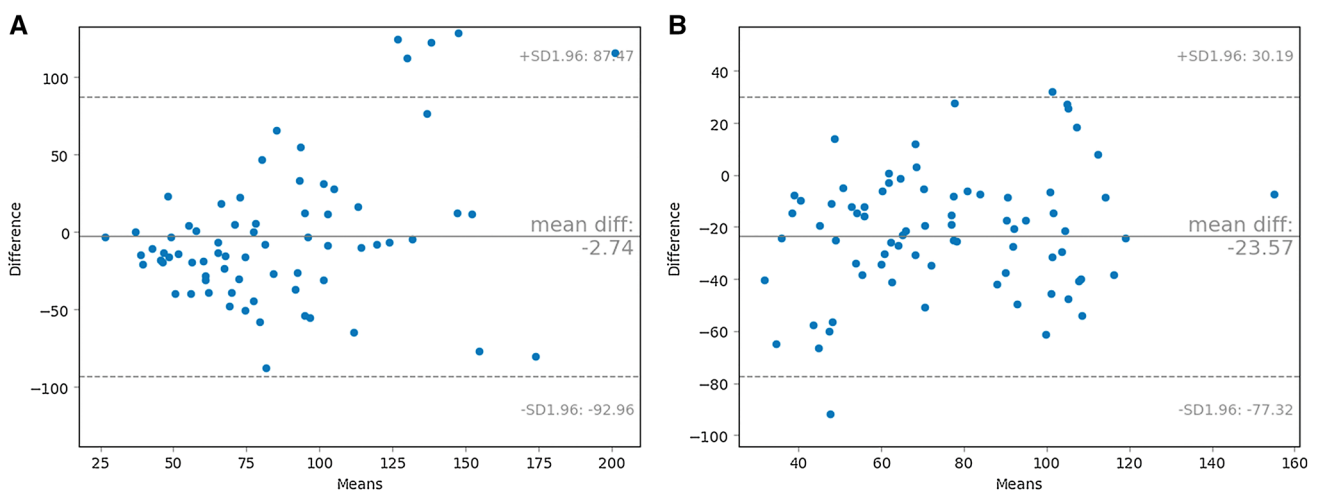
**Table 1** Summary of technologists' performance between the first batch (before supplementary training) and the second batch (after supplementary training) for the cases that needed revision

Performance metrics	First batch	Second batch	<i>p</i> value
DSC (mean $\pm$ SD)	$0.63 \pm 0.15$	$0.63 \pm 0.16$	0.61
JC (mean $\pm$ SD)	$0.48 \pm 0.15$	$0.48 \pm 0.15$	0.61
FP (mean $\pm$ SD)	$0.29 \pm 0.21$	$0.21 \pm 0.10$	0.07
FN (mean $\pm$ SD)	$0.36 \pm 0.20$	$0.43 \pm 0.19$	0.12

DSC Dice–Sorenson coefficient, JC Jaccard coefficient, FP false positive; FN – false negative

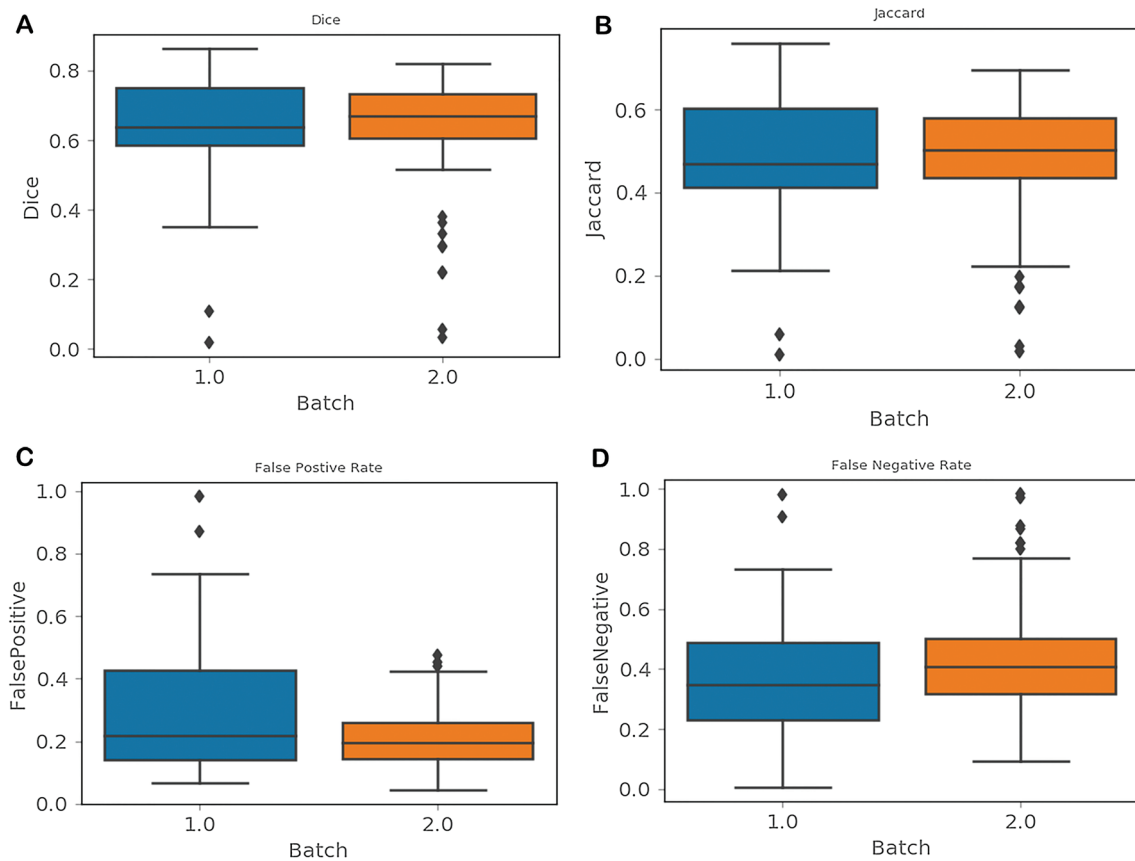
## Discussion

The challenges involved in the curation and labeling of imaging datasets are widely regarded as key barriers for the development and production-scale deployment of reliable AI models in the clinical practice of body imaging. Expert labeling of these datasets is the ideal approach. However, this is often not practical due to the associated costs of time and resources [1]. To the best of our knowledge, training



**Fig. 3** Bland–Altman analyses for mean pancreatic volume difference between technologists' and radiologists' segmentations for cases that required correction before (a) and after supplementary training (b): mean pancreatic volume difference before supplementary training (a)

was  $-2.74$  cc (minimum:  $-92.96$  cc, maximum:  $87.47$  cc). Mean pancreatic volume difference after supplementary training (b) was  $-23.57$  cc (minimum:  $-77.32$  cc, maximum:  $30.19$  cc). Dotted lines indicate limits of differences (mean  $\pm$  1.96 SD)



**Fig. 4** Box and whisker plots of technologists' performance during first (blue, labeled as Batch 1.0) and second batch of segmentations (orange, labeled as Batch 2.0) when compared against radiologist'

segmentations in terms of Dice-Sorenson coefficient (Dice) (a), Jaccard coefficient (Jaccard) (b), false positive rate (c), and false negative rate (d)

technologists for creation of labeled medical imaging datasets have not been explored. In the literature, experiences with crowdsourcing medical imaging tasks to untrained persons in the community-at-large have been described with variable success. Such tasks include annotations of airways, lung nodules, kidney and liver segmentations, and colon polyp classification on CT colonography images [18–21]. Most of these studies concentrate on tasks that require little expertise of the crowd, as the objects to identify either have well-defined geometry or can be easily separated from the background. A similar approach for pancreas segmentation has not been attempted, which is likely due to the complex morphology and geometry of the pancreas. Thus, there is an unmet need for alternate approaches to generate labeled datasets for body imaging AI applications. In this study, we explored the feasibility of training radiology technologists for the development of a CT dataset of volumetric pancreas segmentation for AI applications. Specifically, we evaluated their performance vis-à-vis radiologists after initial training and assessed the impact of supplementary training on their performance for volumetric pancreas segmentation.

Pancreas morphometrics and radiomics are emerging as biomarkers in both endocrine and exocrine disorders of the pancreas [22]. Accurate pancreas segmentation is essential for further investigation and validation of these biomarkers [8, 22]. A manual approach to pancreas segmentation is cumbersome, inaccurate, and not scalable. Therefore, validated methods for automated segmentation of pancreas in clinical practice are necessary. Automated pancreas segmentation will also have potential applications in surgical and radiation therapy planning, and for early detection of pancreatic cancer [5]. Although technologists gain a working knowledge of key anatomical landmarks during their routine clinical assignments, the skills needed for fine segmentations of organs such as pancreas on cross-sectional imaging are not part of their portfolio. Therefore, in this project, we created an image-rich training curriculum focused on multiplanar pancreatic anatomy on CT, which also included common anatomic variations and relevant CT artifacts. Secondly, we conducted instructional tutorials through multiple videoconferencing sessions for the technologists. All of these sessions had been recorded so that future training could be delivered

through as videos or online modules without direct participation by radiologists.

After the initial training, 62% of pancreatic segmentations by the technologists were deemed accurate when compared against the ground truth segmentations by radiologists. Given the inherent complexity of pancreas segmentation, we believe this is an encouraging result that justifies the upfront investment of our time and resources in their training. Secondly, the majority of the errors were due to undersegmentation of pancreatic anatomy. A higher proportion of undersegmentation errors suggests that the technologists generally adopted a cautious approach to the segmentation task, which often augurs well for beginners. The performance of technologists should also be viewed in the context of certain other factors. We did not categorize errors into minor and major classes. Any segmentation that was not deemed accurate was redone. Participation in this project was on a voluntary basis. All the segmentations had to be done during the course of a regular clinical assignment. Although the clinical volumes were low due to the COVID-19 containment phase, the tasks of segmentations were not entirely uninterrupted. We also did not structure additional compensation or time-off into this project. In the future, performance-based rewards and, possibly, gamification of segmentation tasks could augment motivation and performance, as has been observed by others [23–25]. It is also possible that some technologists may not need any subject matter training and could perform reasonably with just instructions on the use of segmentation software and workflow. Secondly, since trainees such as medical students, residents and fellows are also often motivated to participate in medical imaging AI projects, a future prospect is comparison of performance of untrained or trained technologists with that of those trainees, which we plan to undertake in the next phase.

Another important consideration is the software platform used for segmentation tasks. The ground truth pancreatic segmentations were done by radiologists with an AI-assisted segmentation module on 3D Slicer®. This software has to be downloaded on each computer for a given user and requires a certain amount of practice. On the other hand, the technologists used our enterprise custom image-viewing software for their segmentations. This was not a deliberate measure but rather a decision that had to be made in view of the accessibility and their familiarity with the enterprise image-viewing software. This enterprise software is pre-installed on all computers in our institution. Since technologists routinely used this software for their clinical functions, they were well-versed with its basic functions (e.g., loading a study, selecting a particular series, etc.) though they were not aware of its segmentation capabilities. Therefore, our training curriculum and modules included stepwise instructions of the segmentation workflow. This segmentation workflow required the technologists to draw manual regions-of-interest

around the pancreas on each slice. This workflow likely made the segmentations cumbersome, which could have also contributed to the observed errors. Our experience highlights the need for cloud-based image annotation platforms with an intuitive interface that can be seamlessly integrated into the routine imaging workflows.

After the supplementary training, there was a decrease in the range of mean pancreatic volume difference (minimum – 92.96 cc, maximum 87.47 cc in first batch; minimum – 77.32 cc, maximum 30.19 cc in second batch). However, the proportion of accurate segmentations declined to 52%, though the difference against the first batch was not significant. There was also no difference in the similarity metrics in the two batches. Interestingly, the trend towards a decline in segmentation accuracy was primarily due to an increase in the share of undersegmentation errors (63% in first batch and 84% in second batch,  $p=0.003$ ). Conversely, oversegmentation errors significantly reduced (37% in the first batch and 16% in the second batch,  $p=0.003$ ). The decline in oversegmentation suggests that supplementary training helped to better distinguish pancreatic anatomy from subadjacent iso-attenuating structures. However, they likely overcompensated for errors by undersegmenting pancreas at its interface with other organs. Accurate delineation of pancreas margins in areas such as near the duodenal groove can be a challenge even for radiologists. Secondly, our training material and approach could have been inadequate. In the future, improved training modules, more frequent training sessions, assessments over a longer period, and, possibly, a more individualized training approach could result in incremental performance improvement.

It may not be reasonable to expect that technologists' segmentations or labels could be surrogates for that by radiologists. Instead, trained technologists could increase the efficiency of image annotation projects by creating weak labels, which could be used for weakly supervised learning or could subsequently be improvised upon by radiologists [26]. Trained technologists could also augment project pipelines through a review and revision of annotations initially performed by trained AI models. Finally, such a trained group of technologists can be redeployed towards the development of institutional body imaging datasets during both routine instances of scanner downtimes and during extraordinary decline in clinical imaging volumes as was experienced by our institution during our voluntary COVID-19 containment phase.

Our project had limitations. The number and composition of CT scans for this project were based on the ready availability of a curated dataset rather than on statistical considerations. The duration of both initial and supplementary training was relatively short. We also evaluated results for all technologists as a group and could not assess the impact of training on individual performance. We were also unable

to capture the time taken per segmentation because these segmentations had been done during the course of the clinical assignment rather than in controlled research settings.

In summary, trained technologists had a good performance for volumetric pancreas segmentation on CT scans despite complexity of the segmentation task and justified our upfront investment in their training. Such trained technologists could provide a viable option for the development of labeled datasets for body imaging AI applications. Alternatively, they could augment efforts of body radiologists in such development endeavors. The logistics of their engagement will be determined by a given institution's preferences and dynamics of the workplace. There is a need for cloud-based image annotation platforms, validated curriculums, and structured training modules to fully realize the potential of technologists for annotation tasks on body cross-sectional imaging. Investment into these resources could yield a trained workforce that could be gainfully redeployed during routine downtimes as well as during extraordinary circumstances such as COVID-19 containment phase.

**Acknowledgements** The authors thank the CT/MRI technologists at Mayo Clinic, Rochester for their work and participation in this project, and Ishan Garg, MBBS, for his assistance in creation of training materials.

**Author contributions** The authors declare that they have contributed to the design of the work, writing of the manuscript, approval of the final version of the manuscript, and are accountable for the manuscript's contents.

**Funding** Dr. Goenka acknowledges research grant from the Champions for Hope Pancreatic Cancer Research Program of the Funk-Zitiello Foundation and the Advance the Practice Award from the Department of Radiology, Mayo Clinic, Rochester, Minnesota.

**Data availability** The authors declare that they had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis.

## Compliance with ethical standards

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

## References

- Willeminck MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, Folio LR, Summers RM, Rubin DL, Lungren MP (2020) Preparing Medical Imaging Data for Machine Learning. *Radiology* 295 (1):4-15. <https://doi.org/10.1148/radiol.2020192224>
- Miller DD, Brown EW (2019) How Cognitive Machines Can Augment Medical Imaging. *AJR Am J Roentgenol* 212 (1):9-14. <https://doi.org/10.2214/ajr.18.19914>
- Kagoma YK, Netz RJ, Strain A, Larson DB (2018) Improving and Maintaining Radiologic Technologist Skill Using a Medical Director Partnership and Technologist Coaching Model. *AJR Am J Roentgenol* 211 (5):986-992. <https://doi.org/10.2214/ajr.18.19970>
- Chu LC, Park S, Kawamoto S, Wang Y, Zhou Y, Shen W, Zhu Z, Xia Y, Xie L, Liu F, Yu Q, Fouladi DF, Shayesteh S, Zinreich E, Graves JS, Horton KM, Yuille AL, Hruban RH, Kinzler KW, Vogelstein B, Fishman EK (2019) Application of Deep Learning to Pancreatic Cancer Detection: Lessons Learned From Our Initial Experience. *J Am Coll Radiol* 16 (9 Pt B):1338-1342. <https://doi.org/10.1016/j.jacr.2019.05.034>
- Bagheri MH, Roth H, Kovacs W, Yao J, Farhadi F, Li X, Summers RM (2020) Technical and Clinical Factors Affecting Success Rate of a Deep Learning Method for Pancreas Segmentation on CT. *Acad Radiol* 27 (5):689-695. <https://doi.org/10.1016/j.acra.2019.08.014>
- Oda M, Shimizu N, Karasawa K, Nimura Y, Kitasaka T, Misawa K, Fujiwara M, Rueckert D, Mori K (2016) Regression Forest-Based Atlas Localization and Direction Specific Atlas Generation for Pancreas Segmentation. *Lecture Notes in Computer Science* abs/2005.03345:556-563
- Kipp JP, Olesen SS, Mark EB, Frederiksen LC, Drewes AM, Frokjaer JB (2019) Normal pancreatic volume in adults is influenced by visceral fat, vertebral body width and age. *Abdom Radiol (NY)* 44 (3):958-966. <https://doi.org/10.1007/s00261-018-1793-8>
- DeSouza SV, Singh RG, Yoon HD, Murphy R, Plank LD, Petrov MS (2018) Pancreas volume in health and disease: a systematic review and meta-analysis. *Expert Rev Gastroenterol Hepatol* 12 (8):757-766. <https://doi.org/10.1080/17474124.2018.1496015>
- Lu CQ, Wang YC, Meng XP, Zhao HT, Zeng CH, Xu W, Gao YT, Ju S (2019) Diabetes risk assessment with imaging: a radiomics study of abdominal CT. *Eur Radiol* 29 (5):2233-2242. <https://doi.org/10.1007/s00330-018-5865-5>
- Wong VW, Wong GL, Yeung DK, Abrigo JM, Kong AP, Chan RS, Chim AM, Shen J, Ho CS, Woo J, Chu WC, Chan HL (2014) Fatty pancreas, insulin resistance, and beta-cell function: a population study using fat-water magnetic resonance imaging. *Am J Gastroenterol* 109 (4):589-597. <https://doi.org/10.1038/ajg.2014.1>
- Shinagare AB, Steele E, Braschi-Amirfarzan M, Tirumani SH, Ramaiya NH (2016) Sunitinib-associated Pancreatic Atrophy in Patients with Gastrointestinal Stromal Tumor: A Toxicity with Prognostic Implications Detected at Imaging. *Radiology* 281 (1):140-149. <https://doi.org/10.1148/radiol.2016152547>
- Naidich JJ, Boltyenkov A, Wang JJ, Chusid J, Hughes D, Sanelli PC (2020) Impact of the Coronavirus Disease 2019 (COVID-19) Pandemic on Imaging Case Volumes. *J Am Coll Radiol* 17 (7):865-872. <https://doi.org/10.1016/j.jacr.2020.05.004>
- Sammer MBK, Sher AC, Huisman T, Seghers VJ (2020) Response to the COVID-19 Pandemic: Practical Guide to Rapidly Deploying Home Workstations to Guarantee Radiology Services During Quarantine, Social Distancing, and Stay Home Orders. *AJR Am J Roentgenol*:1-4. <https://doi.org/10.2214/ajr.20.23297>
- Center MIR Clinical Trial Processor (CTP). [http://mirewiki.rsna.org/index.php?title=CTP-The\\_RSNA\\_Clinical\\_Trial\\_Processor](http://mirewiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor). Accessed 6/16/2020
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R (2012) 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* 30 (9):1323-1341. <https://doi.org/10.1016/j.mri.2012.05.001>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825-2830



17. Giavarina D (2015) Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 25 (2):141-151. <https://doi.org/10.11613/bm.2015.015>
18. Cheplygina V, Perez-Rovira A, Kuo W, Tiddens HAWM, de Bruijne M Early Experiences with Crowdsourcing Airway Annotations in Chest CT. In, Cham, 2016. *Deep Learning and Data Labeling for Medical Applications*. Springer International Publishing, pp 209-218
19. Boorboor S, Nadeem S, Park JH, Baker K, Kaufman A (2018) Crowdsourcing lung nodules detection and annotation, vol 10579. *SPIE Medical Imaging*. SPIE,
20. Mehta P, Sandfort V, Gheysens D, Braeckeveld G, Berte J, Summers RM Segmenting The Kidney On CT Scans Via Crowdsourcing. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 8-11 April 2019 2019. pp 829-832. <https://doi.org/10.1109/isbi.2019.8759240>
21. Nguyen TB, Wang S, Anugu V, Rose N, McKenna M, Petrick N, Burns JE, Summers RM (2012) Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography. *Radiology* 262 (3):824-833. <https://doi.org/10.1148/radiol.11110938>
22. Almeida RR, Lo GC, Patino M, Bizzo B, Canellas R, Sahani DV (2018) Advances in Pancreatic CT Imaging. *AJR Am J Roentgenol* 211 (1):52-66. <https://doi.org/10.2214/ajr.17.18665>
23. Balducci F, Buono P (2018) Building a qualified annotation dataset for skin lesion analysis through gamification. *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*
24. Vecchio G, Palazzo S, Giordano D, Rundo F, Spampinato C (2020) MASK-RL: Multiagent Video Object Segmentation Framework Through Reinforcement Learning. *IEEE Transactions on Neural Networks and Learning Systems*:1-13. <https://doi.org/10.1109/tnnls.2019.2963282>
25. Winkel DJ, Brantner P, Lutz J, Korkut S, Linxen S, Heye TJ (2020) Gamification of Electronic Learning in Radiology Education to Improve Diagnostic Confidence and Reduce Error Rates. *AJR Am J Roentgenol* 214 (3):618-623. <https://doi.org/10.2214/ajr.19.22087>
26. Zhou Z-H (2018) A brief introduction to weakly supervised learning. *National Science Review* 5:44-53

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