

RESEARCH

Open Access



Algorithms used in medical image segmentation for 3D printing and how to understand and quantify their performance

Magdalene Fogarasi¹, James C. Coburn^{2,3*} and Beth Ripley⁴

Abstract

Background: 3D printing (3DP) has enabled medical professionals to create patient-specific medical devices to assist in surgical planning. Anatomical models can be generated from patient scans using a wide array of software, but there are limited studies on the geometric variance that is introduced during the digital conversion of images to models. The final accuracy of the 3D printed model is a function of manufacturing hardware quality control and the variability introduced during the multiple digital steps that convert patient scans to a printable format. This study provides a brief summary of common algorithms used for segmentation and refinement. Parameters for each that can introduce geometric variability are also identified. Several metrics for measuring variability between models and validating processes are explored and assessed.

Methods: Using a clinical maxillofacial CT scan of a patient with a tumor of the mandible, four segmentation and refinement workflows were processed using four software packages. Differences in segmentation were calculated using several techniques including volumetric, surface, linear, global, and local measurements.

Results: Visual inspection of print-ready models showed distinct differences in the thickness of the medial wall of the mandible adjacent to the tumor. Volumetric intersections and heatmaps provided useful local metrics of mismatch or variance between models made by different workflows. They also allowed calculations of aggregate percentage agreement and disagreement which provided a global benchmark metric. For the relevant regions of interest (ROIs), statistically significant differences were found in the volume and surface area comparisons for the final mandible and tumor models, as well as between measurements of the nerve central path. As with all clinical use cases, statistically significant results must be weighed against the clinical significance of any deviations found.

Conclusions: Statistically significant geometric variations from differences in segmentation and refinement algorithms can be introduced into patient-specific models. No single metric was able to capture the true accuracy of the final models. However, a combination of global and local measurements provided an understanding of important geometric variations. The clinical implications of each geometric variation is different for each anatomical location and should be evaluated on a case-by-case basis by clinicians familiar with the process. Understanding the basic segmentation and refinement functions of software is essential for sites to create a baseline from which to evaluate their standard workflows, user training, and inter-user variability when using patient-specific models for clinical interventions or decisions.

*Correspondence: James.Coburn@fda.hhs.gov

³ Fischell Department of Bioengineering, University of Maryland, College Park, MD 20742, USA

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords: Patient-specific anatomical model, 3D printing, Geometric variation, Medical image segmentation, Segmentation variation, Algorithms, Hospital, Point of care

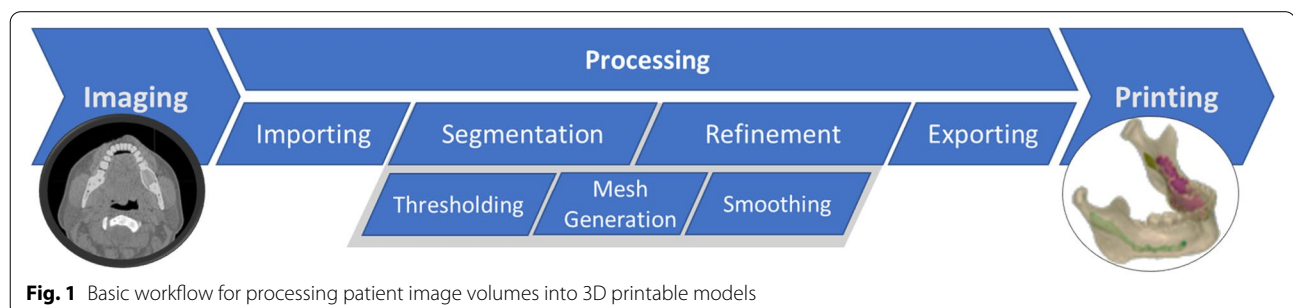
Introduction

Many clinicians are beginning to use 3D printing (3DP), a form of additive manufacturing (AM), to make anatomic models for surgical planning, patient education, and more [1–3]. Over the last decade, traditional manufacturers have used 3DP to fabricate patient-specific medical devices [4] and the breadth of medical image segmentation algorithms has evolved extensively [5–7]. However, the more recent trend is for health care systems to bring 3D printing capabilities within the walls of the hospital at the point of care. With increased implementation at the point of care, health care facilities need to develop methods that ensure these devices are safe and do not increase risk to the patient. In December 2021, the US Food and Drug Administration (FDA) released a discussion paper on the types of 3DP activities projected to be undertaken at the point of care and how they might be regulated [8, 9]. Common use cases include patient-specific implants, surgical cutting guides, and anatomic models [10–13]. Anatomic models may improve surgical outcomes and provide tactile stimulus during surgical planning [14, 15], are the digital base of patient-specific surgical guides, and are the most widespread use case of 3DP in healthcare facilities. Anatomical models may be perceived to have a lower risk compared to cutting guides and implants, but they are still considered medical devices [16] and point of care manufacturers must ensure that these models meet clinical requirements for safety and efficacy. This takes experience and a systematic approach developed from understanding critical attributes of the digital design and fabrication process.

Converting volumetric medical imaging data sets into a 3D printable format involves many steps: a digital pipeline (Fig. 1), where software algorithms are used to isolate anatomic regions of interest (ROI) (referred

to as segmentation), converting the ROI volume to a surface mesh, and cleaning up edges or errors (referred to as smoothing) [17]. Many algorithms and methods have been developed to help users identify boundaries between complex anatomic regions in medical images by automating segmentation and then increasing accuracy through refining processes. These processes can have clinically relevant effects on the final product, depending on the ROI's important features and how the algorithms handle them. Knowledge of key algorithms and parameters can help a user to determine how the software itself will affect the accuracy of the anatomic model and, ultimately, patient safety, irrespective of user variability or other factors.

Many studies have identified absolute differences in accuracy between programs for specific software, [15, 18], described 3D printing workflows for different applications [19–21], and discussed printing best practices [22, 23]. In this paper, the behaviors of different mathematical algorithms used in ROI segmentation and refinement are reviewed to highlight key features that can affect the transformation of medical images into 3D printable models. Next, these concepts are synthesized and illustrated with a real-world clinical case segmented using four different software packages that use various algorithms for segmentation and refinement, including FDA cleared and noncleared software that are either proprietary or open-source. Finally, several metrics are used to quantify local and global geometric variation in 3D models introduced from software workflows. This information provides a framework and tool to assist engineers and clinicians in evaluating and implementing their own processes. The comparison metrics used here can be extended and repeated with other software programs and workflows not described here.



Background: image acquisition and representation of digital anatomy for 3D printing

Volumetric clinical imaging data sets (e.g. computed tomography (CT) and magnetic resonance imaging (MRI)) are typically the source data for 3DP models. These data sets are comprised of volumetric pixels, more commonly referred to as voxels. Slice thickness (z) and pixel size (xy) are a function of scanner hardware parameters, assigned field of view, and matrix size; they affect the overall size and uniformity (isotropy) of voxels. For the same exposure, larger voxels have a higher signal to noise ratio, but at the cost of decreased spatial resolution and increased volume-averaging from adjacent structures co-located within a given voxel [24–28]. Imaging protocols optimized for creation of patient specific 3D models must weigh these trade-offs, and additionally consider contrast requirements, timing, and overall patient positioning; these 3D printing optimized protocols may be different than standard clinical imaging protocols used for diagnosis [4, 29, 30]. Segmentation is the act of assigning voxels that contain an anatomic region of interest (ROI) to a resulting volume or mask. This stair-step appearance is the result of the voxel size and shape and does not represent the true organic contours of tissues or structures. Instead, it contains a portion of adjacent tissue contours when they happen to fall within a given voxel. This stair-step appearance is often softened during the creation of 3D printable surface

meshes from volumetric data. This is because surface meshes are essentially “wrapping” the ROI in a fitted net composed of triangles which can smooth out the blocky voxels.

In this analogy, the net consists of intersections or vertices connected by edges. The space between any three vertices is called a face. (Fig. 2A). Most surfaces begin with uniform size triangles. Increasing the number of triangles in a mesh, much like increasing resolution of the scan, will better approximate curved features. However, dense meshes can be computationally burdensome, so a balance must be struck between complexity and dimensional fidelity. Note that reducing the number of triangles does not always mean that the fidelity will be reduced (Fig. 2B).

Background: Algorithms

Studies of digital workflows found that the largest influences on model accuracy were scan quality and manual segmentation for complex soft tissue cases [17, 26, 30–33]. Recent algorithm improvements for both segmentation and refinement have increased the availability of automated methods [34–36], correspondingly increasing the required user knowledge of algorithms and ability to validate them. The two image processing steps with the greatest chance to impact the final model accuracy are mask editing and mesh smoothing. Once an initial mask is defined, region

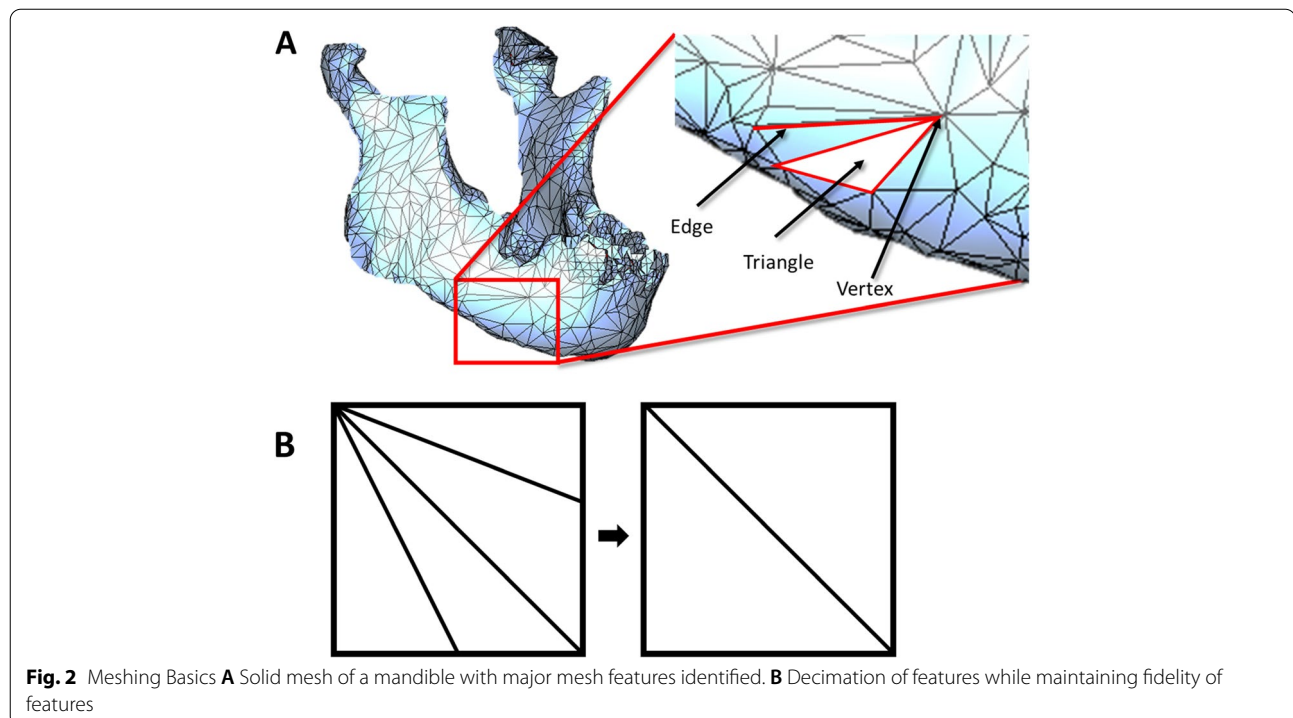


Fig. 2 Meshing Basics **A** Solid mesh of a mandible with major mesh features identified. **B** Decimation of features while maintaining fidelity of features

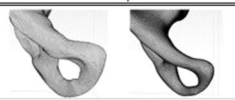
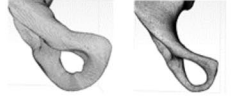
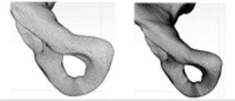



growing, which examines voxels neighboring a seed region, can then be performed to refine the thresholded mask [37]. Programs can then semi-automatically refine segmentation masks using various methods to locate the ROI contours and create the 3D surface mesh. Two examples of these methods are active contours [38, 39] and region competition [40], which both identify and move a contour towards similar nearby voxels while minimizing the contour's deformation. Segmentation results in a volumetric ROI, which is then converted into a surface mesh. The newly created meshes can be smoothed and refined - terms that often overlap. *Refinement* may include any of several methods to increase the fidelity of the mask or the mesh in specific areas where feature resolution is needed. *Smoothing* will refer to algorithms that work on the mesh to decrease quick changes in direction of contours and flatten features according to user defined settings. *Decimation* can reduce the number of triangles that make up a mesh - decreasing file size and complexity while ideally preserving topology [41]. Meshes made from patient images tend to be very complex and decimation reduces computational load.

Most mesh smoothing algorithms function by iterating through the mesh and relocating vertices according to mathematical restrictions that optimize the mesh to a user-set parameter. However, different

smoothing algorithms give different results. This will typically reduce mesh complexity (increase triangle size) in flat areas and increase complexity in areas with many geometric features.

The most common smoothing algorithms implement Laplacian smoothing [42, 43], a vertex-based technique that iteratively converges a curve toward a point. This typically shrinks the volume of the mesh and pure Laplacian implementations do not correct for mesh shrinkage. Modifications such as Taubin smoothing [44] include a compensating inflation step after each mesh shrinkage step. Similar methods such as angle-based [45], bilateral [46], and curvature [47] have been implemented to optimize smoothing in a manner that preserves details (sharp points, small radius curvatures, or thin walls) in the mesh. Most programs include user-selectable options to preserve small features and boundaries. Methods of optimally smoothing a mesh have been an important topic for a long time and they often work extremely well for regular, known or smooth shapes. New implementations and optimizations are constantly being introduced [48–50] to deal with new cases. A summary of popular smoothing algorithms with visual examples is located in Table 1. It is incumbent on the user to identify the characteristics of their software and determine which are most important for their clinical use cases.

Table 1 Summary of most common smoothing algorithms used in segmentation software. Images in Example column demonstrate implementation of the algorithm before (left) and after (right) implementation.

Mesh Smoothing Algorithms				
Smoothing Type	Definition	General Equation	Restrictions	Example
Laplacian	Vertex Based Vertex is relocated to the center of neighboring vertices	Closed Curve Convergence to Point $p_i \leftarrow p_i + \lambda \mathcal{L}(p_i)$ for $0 < \lambda < 1$ $\lambda =$ "damping constant" $p_i^{(t+1)} = p_i^{(t)} + \lambda \Delta p_i^{(t)}$ for meshes	Triangle size effects Without weighting shrinks mesh volume with each iteration May invert edges/triangles	
Recursive Gaussian	Vertex Based Iteratively set each vertex to be linear combination of all neighbors; variant set's vertex to centroid of weighted sum of centroids of adjacent faces	"Zero-Pass" Filtering with adjustments $g(t; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2}$ $\sigma =$ Gaussian $1/2$ width	Too many iterations results in loss of surface details. Choosing too high σ results in no smoothing because of noise/jitter	
Taubin	Vertex Based Laplacian with correction for shrinkage	$p_i \leftarrow p_i + \lambda \Delta p_i$ (shrink) $p_i \leftarrow p_i + \mu \Delta p_i$ (inflate); for $\lambda > 0$ and $\mu < 0$ OR $\mu < 0$ and $ \mu > \lambda$	Requires vertex interpolation to minimize shrinkage	
Angle	Angle Based Assigns minimal angle, iterates through mesh and retriangulates until all of the angles are less than minimal angle.	Node acts like torsion spring $E = \sum_{i=0}^{2(n-1)} \frac{1}{2} k \theta_i^2$ $n =$ # node connected to internal node $k =$ constant $\theta =$ angle between internal and i-th node	Works best on concave domains	
Bilateral	Approximate surface of points via tangent plane and downweight points with poor fit. Based on similarity and closeness	Low Pass Filter Domain $h(x) = k_d^{-1}(x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi) c(\xi, x) d\xi$ Range $h(x) = k_r^{-1}(x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi) s(f(\xi), f(x)) d\xi$ $x =$ neighborhood center $\xi =$ nearby point	Parameters can be selected for edge preservation but are dependent on mesh quality	
Curvature	Mean Curvature Based Adjust local curvature Anisotropic Smoothing	$P_{new} \leftarrow P_{old} + \lambda H(P_{old})n(P_{old})$ $n(P) =$ vector of all inward pointing normals $H(p) =$ amount to move each point inward	Requires calculating principle curvatures to select nodes to shift normals at	

Materials & methods

Materials

The example data set was a cranio-maxillofacial CT scan used in a 2018 RSNA hands-on 3D printing training module [51] (slice thickness 1.0 mm, no gantry tilt, 512 × 512 FOV, 0.3319 mm pixel spacing, 16-bit, 120 kVp, FC80 kernel, and 40 mA). The completely deidentified dataset, available upon request, was provided by Materialise NV (Leuven, Belgium) and was used after receiving approval from the institutional review board. All methods were carried out in accordance with relevant guidelines and regulations. The dataset features a large right mandibular tumor; the tumor is adjacent to both a nerve and an impacted wisdom tooth. This file was chosen because it contains several ROIs with varying degrees of segmentation complexity to challenge different aspects of the digital pipeline. All input images were in Digital Imaging and Communications in Medicine (DICOM) format, and output meshes were saved as STL files.

The four programs selected for this study were Dicom2Print (3D Systems, South Carolina USA), Mimics (Materialise NV; Leuven, Belgium), 3D Slicer (Brigham and Women’s Hospital, Massachusetts USA), and Simpleware (Synposys; California, USA). Throughout, they will be referred to as Programs 1–4 (not ordered as above) because the goal of the study is not to compare the programs themselves. Rather, the goal is to use the programs as examples of different algorithm implementations and to demonstrate the variation that can occur between the workflows of any segmentation software.

All comparisons were made between the final STL surface meshes output from each program. Full mesh comparisons were made using Magics (23.0, Materialise). Linear dimensions and volumetric measurements were computed in 3-Matic (19.0, Materialise). Statistics were calculated using MATLAB (R2007a, Mathworks).









Methods

Digital workflow for STL generation

The process of creating an STL file from DICOM scans follows a similar workflow in all programs, generalized into five main steps: (1) DICOM import, (2) Segmentation, (3) Mesh/Model Generation, (4) Smoothing/Mesh Refinement, and (5) Exporting to STL (or other 3D printer-compatible file format). Initial DICOM imports all maintained the underlying image volume characteristics (e.g., no alterations were made to the source data). As this study was designed to isolate and measure the effects of different algorithms applied throughout the software workflows, a single user (MF) performed all 5 steps of the digital workflow for all conditions in this study (removing any inter-user variability).

The terms automatic, semi-automatic, and manual loosely group interventions by the amount of user intervention required to complete that step of the workflow (Table 2). Regions of interest were segmented in each program, making use of automated and turn-key options when available. For automatic segmentations with default settings, the program-assigned Hounsfield unit range was not adjusted. For semi-automatic thresholding, Hounsfield unit boundaries (minimum and maximum) were set by the user to

Table 2 User intervention level at different workflow steps, by program. Automatic designates use of a menu option or built in automatic “turn key” feature for the majority of the segmentation step. Semi-Automatic designates segmentation steps that required user intervention but made use of built-in operation controls where parameters must be set by the user. Manual designates steps that required all or almost all manual user intervention to segment. Final Models after export were used for subsequent calculations.

Program	Segmentation	Part	Intervention Level				Final Model
			Thresholding	Mesh Generation	Mesh Refinement	Export to STL	
Program 1		Mandible	Automatic	Automatic	Semi-Automatic	Automatic	
		Tumor	Automatic	Automatic	Manual	Automatic	
		Tumor Nerve	Semi-Automatic	Automatic	Manual	Automatic	
		Healthy Nerve	Semi-Automatic	Automatic	Manual	Automatic	
Program 2		Mandible	Automatic	Automatic	Semi-Automatic	Semi-Automatic	
		Tumor	Semi-Automatic	Automatic	Semi-Automatic	Semi-Automatic	
		Tumor Nerve	Semi-Automatic	Automatic	Manual	Semi-Automatic	
		Healthy Nerve	Semi-Automatic	Automatic	Manual	Semi-Automatic	
Program 3		Mandible	Automatic	Automatic	Semi-Automatic	Automatic	
		Tumor	Automatic	Automatic	Manual	Automatic	
		Tumor Nerve	Semi-Automatic	Automatic	Manual	Automatic	
		Healthy Nerve	Semi-Automatic	Automatic	Manual	Automatic	
Program 4		Mandible	Automatic	Automatic	Automatic	Automatic	
		Tumor	Automatic	Automatic	Automatic	Automatic	
		Tumor Nerve	Automatic	Automatic	Semi-Automatic	Automatic	
		Healthy Nerve	Automatic	Automatic	Semi-Automatic	Automatic	

be the same between programs. Global thresholding based on gray values/Hounsfield units was the principle starting point for bone volume segmentation. Thresholding identified the initial mask or created seed markers starting with bone presets (if available), then editing the mask in later steps. Tumor segmentation was performed by using soft tissue CT presets to generate the initial mask and then editing the mask of the tumor area from the mandible using mask splitting. In some of the programs, it was necessary to manually edit the tumor segmentation layer by layer along the bone tissue interface. Segmentation was performed manually for the nerves in all cases by painting on the mask on each layer visible and using interpolation when available to connect the mask layers. Thresholding was attempted for the nerves but because of their resolution it was not an effective method of segmentation. At this point, meshes were generated (step 3 of 5 in the digital pipeline, Fig. 1). This was automatic for all 4 programs.

Mesh refinement (step 4, Fig. 1) used smoothing algorithms that relocated vertices and/or changed the number of triangles (Table 3). Three levels or intensities of smoothing were chosen to best show the changes made by each setting: 0) no smoothing, 1) low smoothing, and 2) high smoothing. Within the semi-automatic intervention level, the amount of user interaction and control varied between programs. In

the final step (step 5, Fig. 1), all meshes were exported as STL files, a condition included in the variability measurements.

Comparison of STL files across levels of smoothing and software programs

A total of 12 models were compared, resulting from three smoothing levels (0, 1, 2) performed in each of 4 programs. Measuring the accuracy of anatomic models often presents a conundrum in that ground truth is only available when scanning cadaveric or simulated models that can be measured separately. Given the inability to perform ground truth verification, we evaluated several metrics for their ability to measure the agreement and disagreement between segmented regions of interest. Volume and surface area measurements provide global comparisons between models generated by each software. Linear measurements provide a means for measuring local differences between models.

Metrics for global comparison

A residual volume comparison available in the literature [52] was used to calculate tumor volumes for all models. These residuals were used to calculate the agreement and disagreement volume and percentages between models (Eqs. 1 and 2), and pairwise statistics were performed. [52, 53]. The agreement (Eq. 1) metric defines the space that is occupied by both models and can be used to assess

Table 3 Smoothing settings for each program given the available smoothing options. Groupings were assigned to generate comparable options from disparate software nomenclature and may not represent exact definitions or labels in their respective software. The volume and surface areas of the three models generated at each smoothing level were averaged and used for statistical analysis.

Modifying Parameters by Program				
	Program 1	Program 2	Program 3	Program 4
Smoothing Options by Program	Type: Laplacian / Taubin Interactions <by number> Smoothing coefficient: Relaxation Factor / Passband	Type: On/Off Interactions <by number> Smoothing coefficient: Smoothing Factor	Type: On/Off Iteration: <by number> Smoothing coefficient: N/A	Type: Automatic Iteration: N/A Smoothing coefficient: 0.01-1mm
Mesh Generation Parameters	Type: Triangle Decimation Modifier: % Decimation Goal	Type: Smoothing Modifier: Low, Medium, High, Optimal, Custom	Type: Triangle Decimation Modifier: N/A	Type: Automatic Modifier: N/A
Model Refinement Settings	Type: Boundary Detection Smoothing Modifier: N/A	Type: Triangle Reduction Modifier: Iterations	Type: Decimation Modifier: Max %	Type: Decimation Modifier: % Reduction

Smoothing Settings by Program					
Smoothing Level	Modifying Parameters	Program 1	Program 2	Program 3	Program 4
0	Smoothing Options by Program	Type: Not Applied Interactions: 0 Smoothing coefficient: N/A	Type: Off Interactions: 0 Smoothing coefficient: N/A	Type: Off Iteration: 0 Smoothing coefficient: N/A	Type: Not Applied Iteration: N/A Smoothing coefficient: N/A
	Mesh Generation Parameters	Type: Triangle Decimation Modifier: 0%	Type: Smoothing Modifier: Optimal	Type: Triangle Decimation Modifier: N/A	Type: Automatic Modifier: N/A
	Model Refinement Settings	Type: Boundary Detection Not Applied Modifier: N/A	Type: Triangle Reduction Modifier: 0	Type: Decimation Modifier: 0 %	Type: Decimation Modifier: 0% Reduction
1	Smoothing Options by Program	Type: Taubin Interactions: 10 Smoothing coefficient: 0.70 Passband	Type: On Interactions: 2 Smoothing coefficient: 0.100	Type: On Iteration: 10 Smoothing coefficient: N/A	Type: Automatic Iteration: N/A Smoothing coefficient: 0.10 mm
	Mesh Generation Parameters	Type: Triangle Decimation Modifier: 70% Decimation Goal	Type: Smoothing Modifier: Custom	Type: Triangle Decimation Modifier: N/A	Type: Automatic Modifier: N/A
	Model Refinement Settings	Type: Boundary Detection Smoothing Modifier: N/A	Type: Triangle Reduction Modifier: Iterations	Type: Decimation Modifier: 60%	Type: Decimation Modifier: 70% Reduction
2	Smoothing Options by Program	Type: Taubin Interactions: 20 Smoothing coefficient: 0.70 Passband	Type: On Interactions: 10 Smoothing coefficient: 0.100	Type: On Iteration: 20 Smoothing coefficient: N/A	Type: Automatic Iteration: N/A Smoothing coefficient: 0.50 mm
	Mesh Generation Parameters	Type: Triangle Decimation Modifier: 70% Decimation Goal	Type: Smoothing Modifier: Custom	Type: Triangle Decimation Modifier: N/A	Type: Automatic Modifier: N/A
	Model Refinement Settings	Type: Boundary Detection Smoothing Modifier: N/A	Type: Triangle Reduction Modifier: Iterations	Type: Decimation Modifier: 60 %	Type: Decimation Modifier: 70% Reduction

accuracy and repeatability in models made by different operators or software programs, or smoothing strategies [54]. Disagreement (Eq. 2) defines the space occupied by only one of the models and not the other.

Surface deviation heatmaps were created for all program pairs by calculating the Hausdorff distance, which is a measure of the distance from all points on one surface to the corresponding points on a second surface, to identify all local areas where variation was more prominent.

Metrics for local comparison

Local comparisons were taken on the final, digital print-ready models output by each program, and allowed for evaluation of software performance on specific ROIs (mandible, tumor, nerve). Linear mandible measurements were performed for all models (across smoothing levels and software programs) with virtual calipers (Magics 23.0) using clinical fiduciary markers [55, 56]. To facilitate linear measurements of the tumor, parallel measurement planes were evenly distributed throughout the tumor region by slicing each STL with identical, parallel datum planes. The centroid of each planar contour was used to measure corresponding X and Y distance to the tumor edges on that contour. Similarly, the alveolar nerve path was measured using the centroid location coordinates of selected coronal slices. Consistent slices were user-selected and evenly distributed along the nerve models.

Statistics

Comparative analytics between the mandible and tumor STL model surface areas, STL model volumes, and nerve path centroids were calculated using ANOVA ($\alpha=0.05$) with Tukey post hoc testing.

Results

Overall observations

The time to process the image volume through all segmentation and refinement steps was comparable for most of the programs, although Program 4 took substantially less time. Manual intervention to some degree was required in all cases but the amount of intervention varied by program. User time to segment in order of program was 4.4 h (Program 1), 4.5 h (Program 2), 4.6 h (Program 3), and 2.0 h (Program 4).

Initial visual inspection of print-ready models showed clear differences in bone contour around the tumor, especially at internal soft tissue boundaries. On the inner left wall of the mandible abutting the tumor, there were noticeable differences in wall thickness, with areas of apparent absent bone as viewed in detail in Fig. 3. Bones with a thin cortical layer are challenging to segment at baseline, and the disruption of the cortex caused by the adjacent tumor compounded the challenge such that most of the programs did not capture all the bone contour along the inner wall.

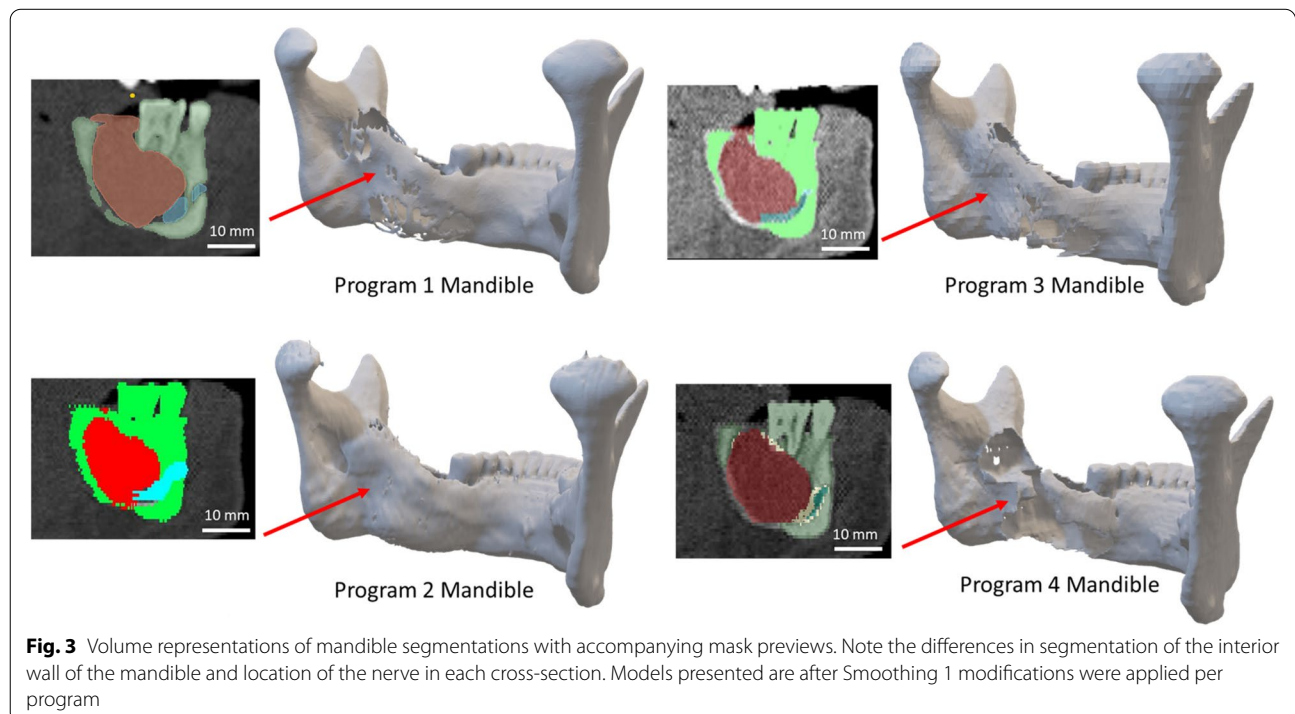


Fig. 3 Volume representations of mandible segmentations with accompanying mask previews. Note the differences in segmentation of the interior wall of the mandible and location of the nerve in each cross-section. Models presented are after Smoothing 1 modifications were applied per program

Effect of different levels of smoothing on final model STL

We tested 12 conditions, including 3 levels of smoothing within each of the 4 programs. The different smoothing levels created no statistically significant differences in final models within each program, therefore only results of program-to-program comparisons are presented. For linear, surface area, and volume measurements, dimensions were averaged across smoothing levels for each program (values for each program represent *average(Smoothing0, Smoothing1, Smoothing2)*). For heatmap and intersection (agreement/disagreement) calculations, smoothing level 1 models were compared between programs.

Quantitative differences using global measurements

Surface deviation heatmaps and residual volume comparisons were used to assess global inter-model differences for models created using smoothing level 2 in each of the four programs. The pair of models with the worst agreement is presented as it demonstrates the best visualization in a static setting (Fig. 4B). Union and intersection figures (Fig. 4A) highlight variations in the outer contour of the tumor between software programs, locating the areas of greatest mismatch. Agreement and disagreement percentages (Table 4) help to generalize this number, with the disagreement metric providing a better sense of the degree of mismatch.

Significant differences ($p < 0.0005$) were found for the averaged final models compared between programs,

for the following parameters: mandible volumes, tumor volumes, and the mandible surface area (compared separately).

Signed difference heat maps of models from each pair of programs illustrated the most problematic locations and the differences seen between segmentation / refinement processes. The pair of models with the largest differences (Fig. 4B) highlighted the areas of high deviation between programs in red (positive deviation) and dark blue (negative deviation). While deviation magnitudes were less than or equal to 2.06 mm even in the worst case, the location of the differences was along the nerve path. Both of these location-specific metrics draws the user’s attention and scrutiny to regions that were determined to be clinically important before the segmentation process began.

Local geometric differences

It is not usually possible to find general metrics that capture all the relevant features of a case in one or two numbers. Additional metrics that can show local model variation and features are essential for validating processes and maintaining output quality based on clinically relevant criteria. They are also often easier to compare with the original DICOM data, as linear measurements are easily performed using PACS software.

Two point-to-point length measurements on the mandible differed with statistical significance using ANOVA ($\alpha = 0.05$) with Tukey post-hoc: Right UCo – Right Go

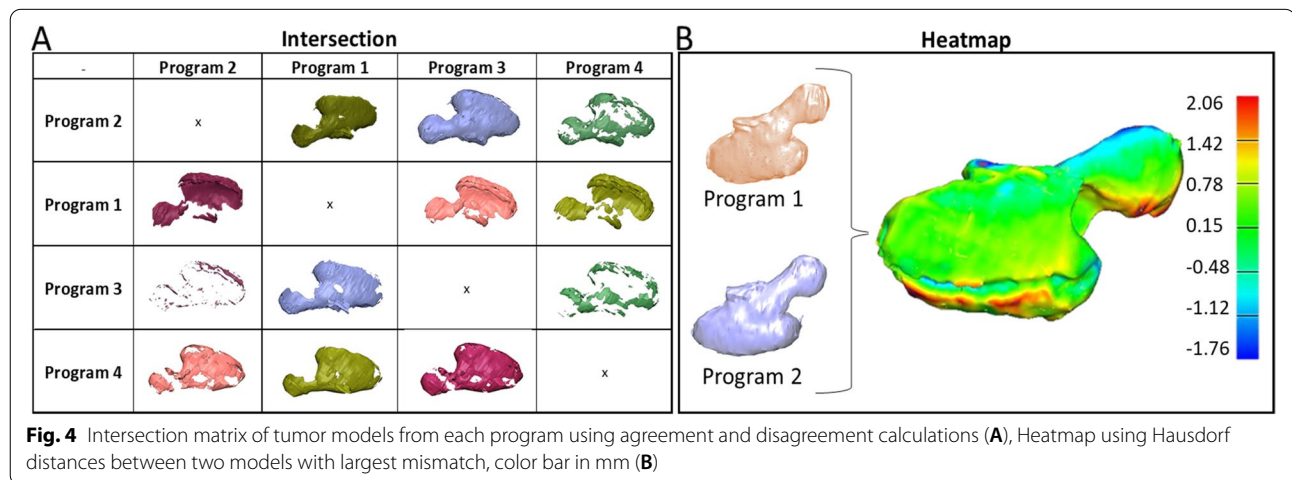
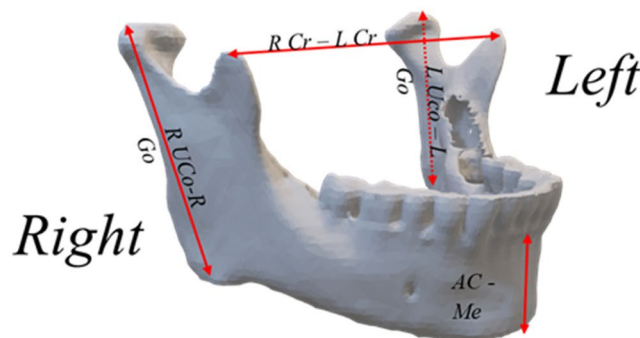


Table 4 Tumor volume agreement and disagreement, in percent (%), rounded to nearest percent.

Calculation	Models					
	Program 1 v 2	Program 1 v 3	Program 1 v 4	Program 2 v 3	Program 2 v 4	Program 3 v 4
Agreement %	81%	82%	82%	94%	94%	93%
Disagreement %	38%	36%	36%	24%	13%	14%



Program	Measurement (mm)			
	Cr-Cr	Right Uco - Right Go	Left Uco - Left Go	Me - AC
Program 1	94.80	62.89	63.13	27.41
Program 2	94.33	61.34*	62.44	27.68
Program 3	93.53	63.08	63.17	27.35
Program 4	94.53	64.34	62.54	25.93*

* Significantly Different from other programs ($p < 0.005$)

Fig. 5 Mandible Linear Measurements. Measurement locations taken (top) and results (bottom). Measurements taken in 3-matic. * ANOVA ($\alpha = 0.5$). Tukey–Kramer $Q_{4,6,0.05} = 4.065$

for Program 2 ($p < 0.005$), and Me-AC for Program 4 ($p < 0.005$) differed from the other programs (Fig. 5) but these values did not appear to be clinically significant.

Local measurements were particularly needed in assessing differences in segmentation of the nerve.

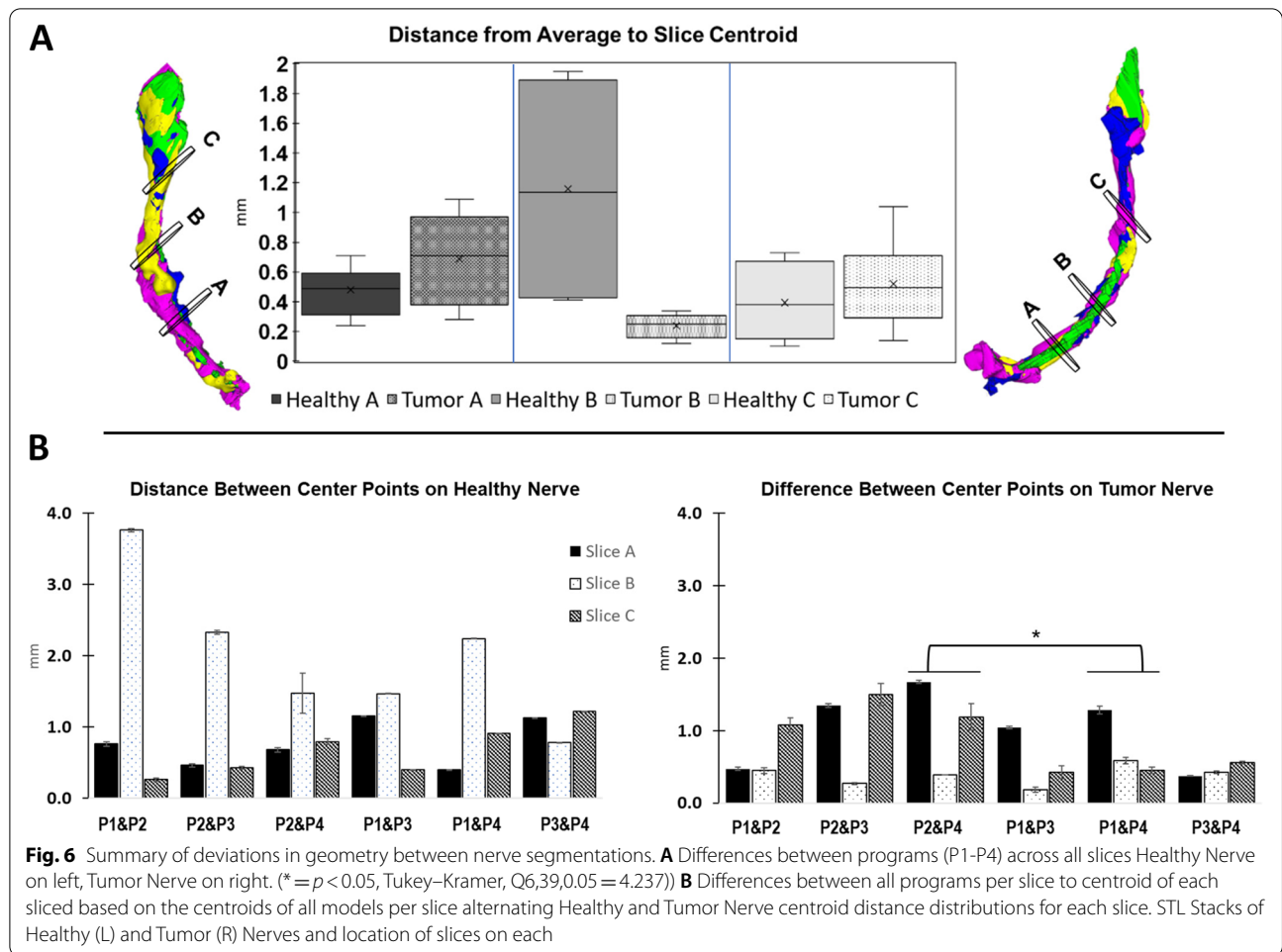
Visual inspection of Intersection metrics and stacked nerve segmented models showed local differences in undulations of the nerve path (Fig. 6A). The average center of each slice was found for both sides and the distance from each centroid to that average was calculated. No significant differences were found in average-to-slice differences on the healthy side, likely due to large variability in the health side distances. Although the centroid location of Slice B varied substantially on the healthy side, there were no statistically significant difference between programs when taking the nerve as a whole. If this side were to be part of the clinical intervention, the individual results would call for closer examination of the nerve path. In contrast, a statistically significant difference was seen in two comparisons (P2&P4 and P1&P4) on the tumor side despite each slice showing differences of less than 2.0 mm (Fig. 6B).

While some statistically significant differences were found for local, linear measurements of tumors, they did not meet the threshold for clinical relevance based on the

criteria for this anatomic ROI and case study. Specifically, the variance in tumor morphology and margins were below the expected clinical margin to be taken when excising the tumor and the location of the nerve did not change in a way that would modify a likely surgical plan. This emphasizes the importance of understanding the clinical relevance of measurements and not solely relying on statistical measures.

Discussion

This study summarized segmentation and refining algorithm features, identified several critical points in data workflows that can introduce variation, and compared methods to measure variation between them when applied to the same data set. No single measurement was the best method for quantifying variation, so it was essential to have both global and local analyses of geometric differences. The intersection volumes and heatmaps were excellent tools to quickly identify the location and size of 3D differences between models. Local linear measurements of the affected nerve provided additional details that were useful in context. Several additional metrics were tried, but the combination of heatmap, volume, and disagreement



metrics provided the most useful information for understanding variability introduced from algorithmic differences in the segmentation / refinement process.

Identifying and quantifying sources of variation is especially important when ground truth measurements cannot be made, as is the case with a patient’s internal anatomy. While cadaveric specimens or polymer models can provide independently verifiable measurements, they often do not fully replicate the living system. Therefore, users need to be able to assess the relative effects of different steps in their workflows to validate that it achieves the clinical needs of the model, while calling out areas where inherent variability will necessitate additional scrutiny. Combining available metrics provided enough local and global data to more holistically understand the variability introduced by different software programs and to identify potentially clinically relevant differences in model output. This study focused particularly on the effects of different software algorithms across programs. In addition to comparing algorithms, the same metrics

can be applied to use of a single program with different settings to determine the effect of those changes (as done here with different levels of smoothing) or to determine the consistency between different users performing established procedures (not done in this study).

As outlined in the background, segmentation, smoothing, and decimation algorithms each have specific effects on the final output. It is critical that software programs preserve anatomical accuracy during the conversion process when these models are used for clinical decisions. Many software packages provide documentation of validation results and activities performed by the developers to verify their accuracy, including but not limited to receiving FDA clearance which reviews this factor. However, if these software packages are used off label for different, untested anatomy, it is incumbent that the user verify that the software provides a quality result. Products without FDA clearance may complete in-house or third-party testing to verify that they are accurate for specific use cases.

In all cases, it is in keeping with both Good Laboratory Practices and Good Manufacturing Practices for each user or institution to test and validate any workflow they use.

Overall, it remains difficult to quantify variations in geometry created with different segmentation and smoothing techniques. Each anatomic ROI also has its own complexities and clinical requirements with regards to accuracy. To assess global differences between models, the volume and disagreement metrics provided quick and very useful overall measures of the match between different processes. In addition, the intersection models could be used to calculate a mismatch volume, surface area, and centroid. These could be particularly useful to monitor inter and intra user variability by setting a benchmark volume and disagreement metric for a standard scan, with users required to meet or beat that benchmark.

It is important to note that each metric has challenges as well. Volumes can only be calculated on closed surfaces, so some regions require additional manual intervention or automated repair before this metric can be calculated. We expected that the surface area metric would be more sensitive than the volume metric due to the irregular shapes in this ROI, however it was too sensitive to be useful. Small differences in the mesh, for example a sharp point or a jagged cluster of triangles, greatly affected the surface area with minimal real effect. In smooth and regular regions, surface area might be a suitable metric with appropriately high sensitivity. It is important to remember that even if a surface or volume metric shows statistically significant differences, those differences are not always clinically relevant and must be assessed according to the clinical needs determined beforehand for that ROI.

The heatmap view was one of the most robust, built-in tools to visualize differences between two models. Many mesh refinement programs include a heatmap generating function. Unlike other tools, they create a visual comparison of the entirety of two surface meshes. However, it is only useful in determining differences in model generation workflows for a given patient model and is not capable of determining the absolute accuracy of the surface mesh compared to the patient scan. Even so, heatmaps are still extremely useful in measuring process quality and repeatability, validating workflows, or assessing whether a new type of clinical case is within the capabilities of the existing protocols. Of note, heatmaps and the disagreement and intersection metrics require that both models be placed exactly in the same location. This is trivial when using a single scan but can be challenging when registering multiple imaging modalities. Model registration can

also sometimes be done using built-in functions after surface meshes are created. However, if they are not available slice centroids and linear measurements may be more robust and effective metrics. The three-dimensional centroid and inertial coordinate system of each mesh can be found as previously described [57, 58].

As expected, regions of high geometric variation occurred in areas of high curvature, in agreement with other studies, [59] most notably in the channel formed by the tumor impinging on the nerve (Fig. 4B). Variation also existed in places susceptible to partial volume averaging, where a feature size was similar to or smaller than a given voxel size. This was illustrated in this use case in regions where there was thin residual cortical bone bordering the tumor. Visual inspection of the DICOM images showed a blurred boundary between the tumor and bone, implying that while the tumor had not breached the bone, it could not be segmented using standard thresholding because of how few voxels the bone occupied in comparison to surrounding soft tissue. When altering segmentation and refinement to include thin walls, care must be taken to maintain the boundaries of other similar tissues in the same area. Automatically processing features like thin walls requires specific quality controls or visual checks to prevent these features from being accidentally removed. While there are many benefits to automated methods, default settings are not always sufficient to preserve small or unexpected features. In many cases, operators must interpret the anatomic regions and modify the segmentation masks manually.

Linear measurements were the easiest way to assess the difference between specific points of interest. Multiple anatomical landmarks were tested for ease of repeatability, leading to the selection of those used in this study. Fiducial markers should be selected for both clinical relevance and repeatability. If an anatomic region has a constant dimension in a critical location for the clinical intervention, then point dimensions are an easy way to assess accuracy (e.g., diameter of a vessel or length from a cut-plane to a nerve). Linear measurements can also be compared to clinical measurements made on the source imaging study, which provides an excellent means of confirming certain key features are accurate between the imaging and the model. It was noteworthy that higher smoothing levels produced larger differences between the models, likely because of the difference in the ways each program controlled for mesh shrinkage.

Limitations

This paper investigates the effects of the implementation of different segmentation and refinement algorithms on the 3D model outputs. To ensure that the effects of

the algorithms were isolated, a single segmenter and single data set were used. This paper cannot compare the magnitude of the algorithmic variation with inter-user or intra-user variability, between different scanning protocols, or in different clinical applications. In clinical use, these variables are comingled and often inextricable. Further studies identifying and using these parameters in the appropriate clinical context will be important to determining the overall effect of these algorithmic variations. However, this paper provides a baseline for understanding the magnitude of algorithmic effects and a set of measurement tools that, when used together, can provide enough information to make decisions as to model quality and process capability.

Risk mitigation

Many studies have focused on dry bone models as their gold standard, but it has been established through this and other studies [60] that partial volume effects related to voxel size are a major cause of segmentation variation

when soft tissue segmentation is necessary. Multiple prior studies have investigated the variations in printed anatomical models [10, 32, 55, 59], however, dry bone, ground truth studies cannot always account for the variability in clinical image volumes. Once a system’s physical accuracy and limits are established through a ground truth study, the accuracy of a patient-matching process typically requires more real-world data.

Minimizing the geometric variation during the DICOM to 3D model conversion process requires understanding of both the software workflow and the clinical needs. It can be thought of as mitigating risk to both the process and to the patient. This is best achieved by using staff trained well in both anatomy and segmentation processes to mitigate both software and inter-user variability. Models should always be assessed against patient data by appropriately trained engineers or clinicians. A checklist of workflow risk mitigations can be used and customized to fit each user’s needs (Fig. 7).

Step	Workflow Check	Reason
Clinical Requirements	<ul style="list-style-type: none"> • Before beginning digital workflow • Identify clinically relevant features and dimensions • Set dimensional tolerance limits for each clinical use 	<ul style="list-style-type: none"> • Establishing success criteria <i>a priori</i> based on objective clinical needs reduces "subjective" changes
DICOM Import	<ul style="list-style-type: none"> • Verify file header information • Check for imaging artifacts (e.g. metal artifact) 	<ul style="list-style-type: none"> • Avoid Import Errors • Identify segmentation compensation or reimage
Segmentation	<ul style="list-style-type: none"> • Provide sufficient training for each use case • Check for and clean up discontinuous regions • Check tissue boundaries in critical regions 	<ul style="list-style-type: none"> • Assess training based on achieving consistent metrics • Floating regions create mesh noise • Extra manual intervention in "grey" areas can greatly improve quality
Mesh Generation	<ul style="list-style-type: none"> • Select minimum triangle size to be at smaller than smallest feature or curve • Check contours in critical or low contrast areas • Check for open edges and mesh discontinuities 	<ul style="list-style-type: none"> • Generates smoother curves and leaves room for later decimation • Critical areas are most likely to show errors • Identifying edges now greatly improves later workflow steps
Smoothing & Refinement	<ul style="list-style-type: none"> • Maintain edges and boundaries • Manually check for voids caused by meshing • Remove duplicate mesh surfaces 	<ul style="list-style-type: none"> • Smoothing can remove key features • Improper meshing can skew metrics • Algorithms can accidentally duplicate surfaces, creating later problems
Exporting	<ul style="list-style-type: none"> • Keep DICOM or PACS coordinate system • Inspecting mesh overlaid on source imaging • Double check saving parameters to maintain all original scaling, origin, and settings 	<ul style="list-style-type: none"> • Consistent coordinates aids comparisons • Manual verification maintains quality control • Avoid future problems by minimizing the changes made by software at this stage

Fig. 7 Check list to mitigate risk in digital workflow process. This is not an exhaustive list but provides basic checks for users to ensure that final STL files exhibit appropriate fidelity to original patient data

Software may be cleared by the FDA for specific applications or intended uses such as segmenting medical images to generate 3D digital models for clinical use. Some of the programs selected for this study have been FDA cleared for specific indications for use (IFU). In that case, the software has shown the FDA that it can create accurate 3D representations of the anatomy listed in its IFU. It also includes sufficient instructions to let trained users replicate those results. While it is not necessary for clinicians to use FDA-cleared software to design anatomic models, if they do not, they would then have to assume the burden of assessing and validating that their software is appropriate and accurate for their use cases. Validation of the software outputs may be done with the assistance of FDA guidance, but algorithm validation is generally left up to the specific vendor. For these reasons, there are benefits to both proprietary software where algorithms have gone through a thorough review process but whose inner workings are hidden, and open-source programs where users can see and control algorithm applications but may lack case-specific validation data. Regardless of the software choice, this study shows that an important aspect of producing models with minimal variability is the user's understanding of the available parameters and how they function within the software of choice. Knowing when to apply and tweak specific parameters is essential to minimizing geometric variability between DICOM sources and final STL models. FDA clearance or third-party validation of software is dependent on the processes used and will not sufficiently mitigate risks unless users understand the algorithms the software applies and how certain parameters may impact the fidelity of the final model.

Clinical relevance

The clinical relevance of geometric deviations in digital models is case dependent and should be assessed by qualified clinicians in reference to the specific procedure and anatomical region of interest. The digital conversion workflow should attempt to maintain the integrity of the original patient imaging data as much as possible. Tumor resection surgeries with larger surgical margins may not be as affected by small surface geometry errors as other surgeries with more stringent requirements. Clinical appropriateness criteria have been developed regarding which cases would benefit from patient-specific models [8].

To minimize the risk for error during digital model creation, it is essential to start with a patient scan of adequate resolution for segmentation. Minimizing voxel size while maintaining a sufficient signal to noise ratio is optimal for achieving high-quality segmentable data. Most diagnostic CT machines are capable of achieving slice thicknesses of less than 1 mm. Most patient-matched

implants are currently accompanied by a specific scan protocol that is optimized for the anatomy around the implant location. Unique cases such as trauma, cancer, and congenital deformity that present some of the best potential for patient-matched technology do not yet have standardized requirements and should be evaluated based on clinical needs.

Conclusions

Quantifying the variations in model design will be essential for patient-matched technology to reach maximum potential. This study provides comparisons of several metrics that can be used to validate methods of preparing patient-matched 3D prints. The software packages used here are representative of many that are available, all of which use similar mathematical foundations to identify regions of interest, then refine the regions to provide the best balance of accuracy and complexity. This provides very concordant results for regular and slowly varying smooth shapes. Most of the model disagreement then occurs during the refinement stages. Each program has features that refine models while reducing noise, preserving sharp curves, or maintaining the overall model volume. Any of these aspects may be desired or undesirable depending on the application. Once known, the implication of variation needs to be assessed by a qualified clinician per clinical application and often per case. Using whole-model metrics that can show the locations of variability (e.g. heatmaps and intersections) coupled with local slice-based measured (e.g. nerve slice centroids) balances the amount of measurement effort with information provided. For maintenance of a process, aggregate measures such as model volume and the disagreement metric were effective in identifying when one segmentation/refinement was sufficiently different from another to require additional checks. The algorithms and metrics described here can facilitate comparison of STL model consistency and accuracy regardless of workflow or program.

A basic understanding of the functionality of segmentation software is essential for ensuring patient safety as medical 3DP continues to expand. Methods used in this paper can help 3D printing facilities establish best practices for evaluating variation between segmentation methods and will allow users to develop optimized workflows—ideally accelerating the patient matched instrumentation implementation in industry and at point of care.

Abbreviations

3DP: 3D Printing; AM: Additive Manufacturing; CT: Computed Tomography; DICOM: Digital Imaging and Communications in Medicine; FDA: US Food and Drug Administration; IFU: Indications for Use; MRI: Magnetic Resonance Imaging; ROI: Region of Interest; STL: Standard Tessellation Language.

Acknowledgements

DICOM data set was provided by Materialise NV, Leuven, Belgium from their data library. 3D Printing resources were provided by the FDA Additive Manufacturing for Medical Products Laboratory.

Notes: Commercially available programs for medical segmentation are cleared by the FDA under specific indications for use and are evaluated for accuracy under those constrained use cases. It is prudent for end users to evaluate any FDA cleared software for applications outside of the published intended use.

Authors' contributions

MF analyzed the segmentation data, organized algorithm background information and contributed to writing the manuscript. JC edited the manuscript and contributed to writing. BR contributed clinical perspectives and assisted in editing. All authors read and approved the final manuscript.

Funding

This project was supported in part by an appointment to the Research Participation Program at the Office of Science and Engineering Laboratories Center for Devices and Radiological Health Division of Biomedical Physics, U.S. Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and FDA.

Availability of data and materials

DICOM data were used in a 2018 Radiological Society of North America training course and provided by Materialise NV. They are available upon request medical@materialise.be. Programs used and their specific versions are available from the authors upon request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹US Food and Drug Administration, Center for Device and Radiological Health, Silver Spring, MD 20993, USA. ²US Food and Drug Administration, Office of the Chief Scientist, Silver Spring, MD 20993, USA. ³Fischell Department of Bioengineering, University of Maryland, College Park, MD 20742, USA. ⁴US Department of Veterans Affairs, Veterans Health Administration, Office of Healthcare Innovation and Learning, Seattle, WA 98109, USA.

Received: 21 January 2022 Accepted: 30 May 2022

Published online: 24 June 2022

References

- SME. Medical Additive Manufacturing/3D Printing. SME Annual Report. 2018.
- Christensen A, Wake N. Chapter 9 - Regulatory Perspectives for 3D Printing in Hospitals. In: Wake N, editor. *3D Printing for the Radiologist*: Elsevier. 2022;109–16.
- Bücking TM, Hill ER, Robertson JL, Maneas E, Plumb AA, Nikitichev DI. From medical imaging data to 3D printed anatomical models. *PLoS ONE*. 2017;12(5):e0178540.
- SME. Physicians as Manufacturers: The Rise of Point-of-care Manufacturing SME Report. 2018.
- Ma Z, Tavares JMR, Jorge RN, editors. A review on the current segmentation algorithms for medical images. Proceedings of the 1st international conference on imaging theory and applications (IMAGAPP); 2009.
- Ramesh K, Kumar GK, Swapna K, Datta D, Rajest SS. A review of medical image segmentation algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*. 2021;7(27): e6.
- Zhang YJ, editor A review of recent evaluation methods for image segmentation. Proceedings of the Sixth International Symposium on Signal Processing and its Applications (Cat No 01EX467). 2001: IEEE.
- Chepelev L, Wake N, Ryan J, Althobaity W, Gupta A, Arribas E, et al. Radiological Society of North America (RSNA) 3D printing Special Interest Group (SIG): guidelines for medical 3D printing and appropriateness for clinical scenarios. 2018;4(1):11.
- FDA/CDRH U. 3D Printing Medical Devices at the Point of Care: Discussion Paper. 2021;1–19.
- Bucking TM, Hill ER, Robertson JL, Maneas E, Plumb AA, Nikitichev DI. From medical imaging data to 3D printed anatomical models. *PLoS ONE*. 2017;12(5): e0178540.
- Nayak UA, Balachandra M, Manjunath K, Kurady R. Validation of Segmented Brain Tumor from MRI Images Using 3D Printing. *Asian Pacific Journal of Cancer Prevention: APJCP*. 2021;22(2):523.
- Liew Y, Beveridge E, Demetriades AK, Hughes MA. 3D printing of patient-specific anatomy: a tool to improve patient consent and enhance imaging interpretation by trainees. *Br J Neurosurg*. 2015;29(5):712–4.
- Heller M, Bauer H-K, Goetze E, Gielisch M, Roth KE, Drees P, et al. Applications of patient-specific 3D printing in medicine. *Int J Comput Dent*. 2016;19(4):323–39.
- Punyaratabandhu T, Liacouras PC, Pairojboriboon S. Using 3D models in orthopedic oncology: presenting personalized advantages in surgical planning and intraoperative outcomes. *3D printing in medicine*. 2018;4(1):12.
- Chae MP, Chung RD, Smith JA, Hunter-Smith DJ, Rozen WM. The accuracy of clinical 3D printing in reconstructive surgery: literature review and in vivo validation study. *Gland Surg*. 2021;10(7):2293.
- Administration UFaD. 3D Printing Medical Devices at the Point of Care: Discussion Paper FDA.gov2021 [1:Available from: <https://www.fda.gov/medical-devices/3d-printing-medical-devices/3d-printing-medical-devices-point-care-discussion-paper>.
- Byrne N, Velasco Forte M, Tandon A, Valverde I, Hussain T. A systematic review of image segmentation methodology, used in the additive manufacture of patient-specific 3D printed models of the cardiovascular system. *JRSM Cardiovasc Dis*. 2016;5:2048004016645467.
- Wijnen N, Brouwers L, Jebbink EG, Heyligers JM, Bemelman M. Comparison of segmentation software packages for in-hospital 3D print workflow. *Journal of Medical Imaging*. 2021;8(3): 034004.
- Bertolini M, Rossoni M, Colombo G. Operative Workflow from CT to 3D Printing of the Heart: Opportunities and Challenges. *Bioengineering*. 2021;8(10):130.
- Gharleghi R, Dessalles CA, Lal R, McCraith S, Sarathy K, Jepson N, et al. 3D Printing for Cardiovascular Applications: From End-to-End Processes to Emerging Developments. *Ann Biomed Eng*. 2021;49(7):1598–618.
- Sharma A, Kirkland KS, Holloway RM, Parekh SG. Incorporating 3D printing into your practice: lessons learned. *Foot Ankle Spec*. 2021;14(4):361–5.
- Piedra-Cascón W, Krishnamurthy VR, Att W, Revilla-León M. 3D printing parameters, supporting structures, slicing, and post-processing procedures of vat-polymerization additive manufacturing technologies: A narrative review. *J Dent*. 2021;109: 103630.
- Bastawros S, Wu L, Liacouras PC, Levin DB, Ahmed MT, Strzelecki B, et al. Establishing 3D printing at the point of care: basic principles and tools for success. *Radiographics*. 2022;42(2):451–68.
- Ibrahim N, Parsa A, Hassan B, van der Stelt P, Aartman IH, Wismeijer D. The effect of scan parameters on cone beam CT trabecular bone microstructural measurements of the human mandible. *Dentomaxillofac Radiol*. 2013;42(10):20130206.
- Whymys BJ, Vorperian HK, Gentry LR, Schimek EM, Bersu ET, Chung MK. The effect of computed tomographic scanner parameters and 3-dimensional volume rendering techniques on the accuracy of linear, angular, and volumetric measurements of the mandible. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2013;115(5):682–91.
- Huottilainen E, Paloheimo M, Salmi M, Paloheimo K-S, Björkstrand R, Tuomi J, et al. Imaging requirements for medical applications of additive manufacturing. *Acta Radiol*. 2014;55(1):78–85.
- Mitsouras D, Liacouras PC, Wake N, Rybicki FJ. RadioGraphics update: medical 3D printing for the radiologist. *Radiographics*. 2020;40(4):E21–3.

28. Furlow B. Medical 3-D printing. *Radiologic technology*. 2017;88(5):519CT-37CT.
29. Christensen A, Rybicki FJ. Maintaining safety and efficacy for 3D printing in medicine. *3D printing in medicine*. 2017;3(1):1–10.
30. Mitsouras D, Liacouras P, Imanzadeh A, Giannopoulos AA, Cai T, Kumamaru KK, et al. Medical 3D printing for the radiologist. *2015*;35(7):1965–88.
31. van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Medical engineering physics*. 2018;51:6–16.
32. Huotilainen E, Jaanimets R, Valášek J, Marcián P, Salmi M, Tuomi J, et al. Inaccuracies in additive manufactured medical skull models caused by the DICOM to STL conversion process. *Journal of Cranio-Maxillofacial Surgery*. 2014;42(5):e259–65.
33. Fourie Z, Damstra J, Schepers RH, Gerrits PO, Ren Y. Segmentation process significantly influences the accuracy of 3D surface models derived from cone beam computed tomography. *Eur J Radiol*. 2012;81(4):e524–30.
34. Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. *Sci Rep*. 2020;10(1):13724.
35. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Seminars in Radiation Oncology*. 2019;29(3):185–97.
36. Gallivanone F, Interlenghi M, Canervari C, Castiglioni I. A fully automatic, threshold-based segmentation method for the estimation of the metabolic tumor volume from PET images: validation on 3D printed anthropomorphic oncological lesions. *J Instrum*. 2016;11(01):C01022.
37. Withey DJ, Koles ZJ, editors. *Medical image segmentation: Methods and software*. 2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging. 2007;IEEE.
38. Caselles V, Catté F, Coll T, Dibos FJNM. A geometric model for active contours in image processing. *1993*;66(1):1–31.
39. Caselles V, Kimmel R, Sapiro GJJocv. Geodesic active contours. *1997*;22(1):61–79.
40. Zhu SC, Yuille AJToPA, Intelligence M. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *1996*;9:884–900.
41. Botsch M, Pauly M, Kobbelt L, Alliez P, Lévy B, Bischoff S, et al. *Geometric modeling based on polygonal meshes*. 2007.
42. Bade R, Jens Haase, Preim B. Comparison of Fundamental Mesh Smoothing. *SimVis*. 2006;6:289–304.
43. Field DA. Laplacian smoothing and Delaunay triangulations. *Communications in applied numerical methods*. 1988;4(6):709–12.
44. Taubin G. A Signal Processing Approach To Fair Surface Design. *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995:351–8.
45. Zhou T, Shimada KJL. An Angle-Based Approach to Two-Dimensional Mesh Smoothing. *IMR*. 2000;2000:373–84.
46. Tomasi C, Manduchi R, editors. *Bilateral filtering for gray and color images*. Iccv. 1998.
47. Semenova I, Savchenko V, Hagiwara IJWToS. Curvature based mesh improvement. *WSEAS Transactions on Systems*. 2004;3(4):1346–54.
48. Freitag LA. On combining Laplacian and optimization-based mesh smoothing techniques. Argonne National Lab., IL (United States). 1997.
49. Bade R, Haase J, Preim B, editors. *Comparison of Fundamental Mesh Smoothing Algorithms for Medical Surface Models*. *SimVis*. 2006;Citeseer.
50. Wei M, Wang J, Guo X, Wu H, Xie H, Wang FL, et al. Learning-based 3D surface optimization from medical image reconstruction. *Opt Lasers Eng*. 2018;103:110–8.
51. Wake N, Alexander AE, Christensen AM, Liacouras PC, Schickel M, Pietila T, et al. Creating patient-specific anatomical models for 3D printing and AR/VR: a supplement for the 2018 Radiological Society of North America (RSNA) hands-on course. *2019*;5(1):1–10.
52. George E, Liacouras P, Rybicki FJ, Mitsouras D. Measuring and Establishing the Accuracy and Reproducibility of 3D Printed Medical Models. *RadioGraphics*. 2017;37(5):1424–50.
53. Cai T, Rybicki FJ, Giannopoulos AA, Schultz K, Kumamaru KK, Liacouras P, et al. The residual STL volume as a metric to evaluate accuracy and reproducibility of anatomic models for 3D printing: application in the validation of 3D-printable models of maxillofacial bone from reduced radiation dose CT images. *3D Print Med*. 2015;1(1):2.
54. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. *Acad Radiol*. 2004;11(2):178–89.
55. Ibrahim D, Broilo TL, Heitz C, de Oliveira MG, de Oliveira HW, Nobre SM, et al. Dimensional error of selective laser sintering, three-dimensional printing and PolyJet models in the reproduction of mandibular anatomy. *J Craniomaxillofac Surg*. 2009;37(3):167–73.
56. Olayemi AB. Assessment and determination of human mandibular and dental arch profiles in subjects with lower third molar impaction in Port Harcourt. *Nigeria Ann Maxillofac Surg*. 2011;1(2):126–30.
57. Coburn JC, Upal MA, Crisco JJ. Coordinate systems for the carpal bones of the wrist. *J Biomech*. 2007;40(1):203–9.
58. Miranda DL, Rainbow MJ, Leventhal EL, Crisco JJ, Fleming BC. Automatic determination of anatomical coordinate systems for three-dimensional bone models of the isolated human knee. *J Biomech*. 2010;43(8):1623–6.
59. Pinto JM, Arrieta C, Andia ME, Uribe S, Ramos-Grez J, Vargas A, et al. Sensitivity analysis of geometric errors in additive manufacturing medical models. *Med Eng Phys*. 2015;37(3):328–34.
60. Rendón-Medina MA, Andrade-Delgado L, Telich-Tarriba JE, Fuente-del-Campo A, Altamirano-Arcos CA. Dimensional error in rapid prototyping with open source software and low-cost 3D-printer. *Plast Reconstr Surg Glob Open*. 2018;6(1):e1646.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

