Using Fuzzy Logics to Determine Optimal Oversampling Factor for Voxelizing 3D Surfaces in Radiation Therapy

C Pinter[1], T Olding[2], L J Schreiner[2] and G Fichtinger[1]

[1] Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Goodwin Hall, 25 Union St, Kingston, ON, K7L 2N8, Canada

[2] Department of Physics, Engineering Physics, and Astrophysics, Queen's University, Stirling Hall, 64 Bader Lane, Kingston, ON, K7L 3N6, Canada

Corresponding author: Csaba Pinter

Email: csaba.pinter@queensu.ca

Phone: +1-613-893-2706

Orcid: 0000-0001-9982-3194

Acknowledgements:

The authors have no relevant conflicts of interest to disclose. This work was partially funded by the Ontario Consortium for Adaptive Interventions in Radiation Oncology (OCAIRO), and An Applied Cancer Research Unit of Cancer Care Ontario with funds provided by the Ministry of Health and Long-Term Care. G.F. was funded as a Cancer Care Ontario Research Chair.

Abstract. Voxelizing three-dimensional surfaces into binary image volumes is a frequently performed operation in medical applications. In radiation therapy (RT), dose-volume histograms (DVH) calculated within such surfaces are used to assess the quality of an RT treatment plan in both clinical and research settings. To calculate a DVH, the 3D surfaces need to be voxelized into binary volumes. The voxelization parameters may considerably influence the output DVH. An effective way to improve the quality of the voxelized volume (*i.e.* increasing similarity between that and the original structure) is to apply oversampling to increase the resolution of the output binary volume. However, increasing the oversampling factor raises computational and storage demand. This paper introduces a fuzzy inference system that determines an optimal oversampling factor based on relative structure size and complexity, finding the balance between voxelization accuracy and computation time. The proposed algorithm was used to automatically calculate oversampling factor in four RT studies, two phantoms and two real patients. The results show that the method is able to find the optimal oversampling factor in most cases, and the calculated DVHs show good match to those calculated using manual overall oversampling of two. The algorithm can potentially be adopted by RT treatment planning systems based on the open-source implementation to maintain high DVH quality, enabling the planning system to find the optimal treatment plan faster and more reliably.

Keywords: Voxelization, Fuzzy logics, Radiation therapy, Dose volume histogram.

1. Introduction

Radiation therapy (RT) is a form of cancer treatment, which uses high-energy radiation to kill cancer cells by damaging their DNA. However, the radiation can also kill healthy cells, so carefully designed radiation plans need to be designed that targeting mainly the cancerous cells. A basic tool for evaluating radiation therapy (RT) treatment plan quality is the dose-volume histogram (DVH) that provides dose information for target and organ-at-risk structures that is easily assessable by the radiation treatment care providers. As DVH and derived metrics often form the objective function in automatic treatment plan optimization, their accuracy is important to ensure that indeed the most optimal plan is selected. The task group commissioned by the American Association of Physicists in Medicine to identify features to include in RT quality assurance considered DVH one of such features to be analyzed (Fraass *et al.* 1998). The role and importance of DVH in RT treatment planning have been shown by several research groups as well (Nelms *et al.* 2015, Corbett *et al.* 2002, Sunderland *et al.* 2016).

An early step in DVH calculation is converting the structures that are stored in series of 2D planar contours into 3D binary volumes, i.e. voxelization. These planar contours are often converted into 3D surfaces as a pre-processing step so that voxelization can be performed more easily. Although the voxelization step itself is rather straightforward, there are certain decisions to be made that may significantly influence the resulting DVH. One such decision is whether to include each border voxel, *i.e.* voxels only a part of which falls inside the structure. This is an especially important factor in areas of high dose gradient (often at the edges of organs-at-risk in an effort to spare the structure), where inclusion or exclusion of a single border voxel may influence the DVH in such a way that it compromises the treatment plan optimization process, potentially resulting in a sub-optimal plan being selected by the planning system (Leasure *et al.* 2017).

Another decision to be made is the resolution of the grid that is used for voxelization. The accompanying dose distribution volume is typically used as a grid reference, but its resolution is unsatisfactory in case of smaller or more complex structures, where a high ratio of the voxels are border voxels. Using a too low resolution voxelization grid can introduce several types of artifacts in the DVH, potentially distorting DVH metrics used in plan optimization. These artifacts are most likely to appear in high dose gradient areas, which means that the targets and the most endangered organs at risk are most prone to produce inaccurate metrics, potentially steering the plan optimizer away from the real optimum. Figure 1.a shows the “staircase” artifact, caused by the low number of voxels taking part in the calculation, and Figure 1.b shows consistent shift, which is caused by a difference in coverage at the borders. As the grid resolution directly influences the proportion of border voxels, which we established above as critical regarding the measurements, by increasing the grid resolution, the DVH inaccuracies can be mitigated. In this situation, dividing the reference voxels into smaller equally sized voxels (i.e. oversampling) seems to be a good idea, because then the borders of small or complex structures can be approximated more accurately.

The number of divisions along each axis is called the oversampling factor. Simply increasing the oversampling factor for each structure, however, is not feasible, because it poses a greater computational and storage demand, which may increase the processing time beyond the clinically acceptable range (time is important because thousands of DVHs are calculated in each plan optimization process). To mitigate this issue, determination of the oversampling factor for each structure is necessary. Application of such per-structure voxelization grid resolution was also suggested by Drzymala *et al.* (1991) in their overview paper about DVHs. Although manual selection of these factors is a possibility, it would not be feasible due to i) the added unnecessary complication to the planning process, ii) the required additional staff training, and iii) its time requirements beyond clinical applicability.

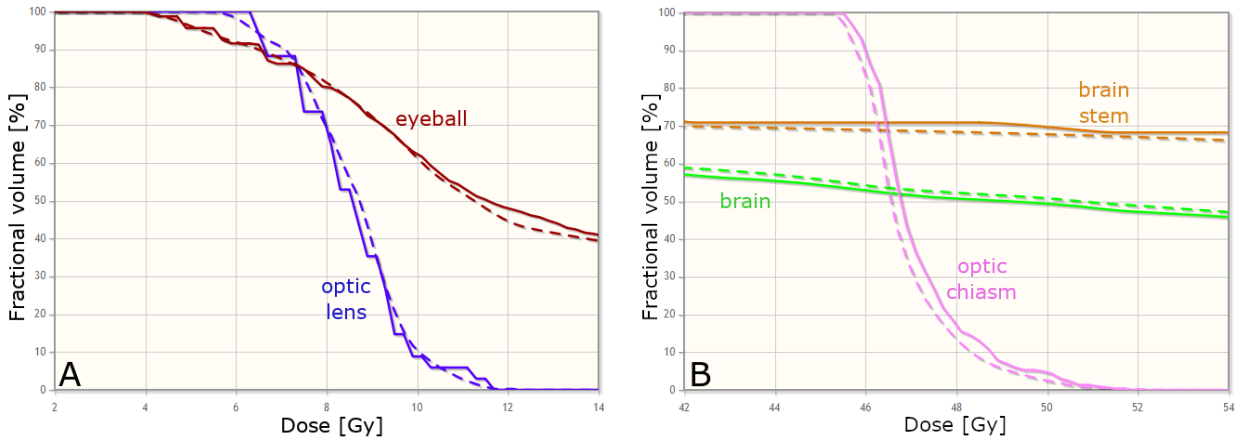

Fig. 1 Representative DVHs from a head and neck analytical dataset to illustrate the potential artifacts associated with voxelization of structures at low resolution (2.5mm isotropic, typical dose volume spacing in clinical setting – solid line) and high resolution (0.25mm isotropic – dashed line). The resolution effects are more pronounced in small sized structures than of large structures. A: Staircase effect on histogram of an optic lens (blue lines) and eyeball (maroon lines), caused by the low number of voxels (samples) available for that structure. B: Consistent shift on the histogram of an optic chiasm (pink lines), brain stem (brown), and brain (green), caused by the difference of coverage of the structure's border on the dose volume

Extensive research has been performed on the properties of DVHs in terms of their usability in radiation oncology. The most active research area is the aforementioned treatment plan optimization, where DVHs are used to evaluate potential plans to automatically find the optimal plan (*e.g.* Kirisits *et al.* 2005; Babier *et al.* 2018, Mountris *et al.* 2019). DVHs are often used to compare treatment techniques (*e.g.* Das *et al.* 2008; Hazell *et al.* 2016, Makishima *et al.* 2019). Its utility in outcome prediction has been assessed in numerous publications as well, as summarized by Rodrigues *et al.* (2004) and Rose *et al.* (2009). Limited literature exists, however, about the accuracy of DVHs in terms of various structures and dose grid resolutions: the most relevant publications are the original paper about DVHs by Drzymala *et al.* (1991), and the article by Nelms *et al.* (2015), the validation dataset from which is used to evaluate the proposed method in section 3. Notable research in the area includes the method Kooy *et al.* (1993) suggesting a non-uniform sampling formalism, the article of Corbett *et al.* (2002) investigating the effect of voxel size on DVH accuracy for a specific procedure, and the paper of Sunderland *et al.* (2016), which can be considered the precursor of this research.

The topic of voxelization has also been investigated in depth. Efficient methods have been proposed using graphics cards to maximize performance (*e.g.* Eisemann *et al.*, 2008; Schwarz *et al.*, 2010). Specialized algorithms have been proposed to enforce particular geometric properties, such as topology (Huang *et al.*, 2014) or sharp edges (Novotný *et al.*, 2010). To our knowledge, however, no publication discusses accuracy in terms of voxelization grid resolution, may it be constant or variable.

The objective of this research is to automate the selection of the optimal oversampling factor to avoid both inaccurate voxelizations potentially resulting in erroneous treatment plan optimization and clinically unfeasible computation time. We propose a fuzzy-based method for calculation of the oversampling factor, so that the DVH calculation process only uses high oversampling factors in case of the aforementioned problematic small or complex structures, but uses the standard reference grid (factor of one) for the structures that are relatively insensitive to the grid resolution, or even subsamples the volume for especially large structures (factor of one half).

Specifically, this research contributes to data science innovation in the following ways:

- Introduce the concept of per-structure voxelization grid (determined by the oversampling factor), optimized for the relevant geometric properties of volumetric structures
- Per-structure voxelization parameters allow accurate quantification of metrics for each structure, while keeping the computational burden low
- Introduce a fuzzy inference system for calculating the per-structure oversampling factor

The hypothesis is that the proposed method finds the optimal oversampling factor in various meaningful test cases. It aims to increase DVH quality in a clinically feasible way, while leaving the planning workflow unchanged. The proposed system and its details are presented in section 2. Evaluation of accuracy and speed are discussed in section 3: the calculated results are compared against baselines determined by experts in four RT studies, followed by validation against a standard synthetic dataset, then against commercial and research software. Summary of the results and future work are discussed in section 4.

2. Methods

A few factors may influence the choice of the magnitude of oversampling, the two most significant of which are relative structure size and complexity. The smaller the structure, the greater the ratio of the border voxels among all structure voxels, which means greater uncertainty of volume measurement and dose value inclusion (or exclusion). In these cases we can increase the accuracy by increasing the oversampling factor. In the case of large structures the opposite is true, so using low oversampling values can be used to reduce computation time, while not sacrificing accuracy. In this work we measure accuracy based on comparison of DVHs calculated on the structure images voxelized using the proposed method, and DVHs calculated by third party applications. This is necessary because commercial treatment planning systems that serve as validation baseline do not give access to intermediate data such as voxelized structure images. The other deciding factor is structure complexity, as it also influences the ratio of border

voxels. If a structure is simple, then most of the included voxels fall inside the volume, while if it is complex, then the border voxels are more emphasized.

It can be noted that we have used terms such as “relatively small”, and “complex” which may be straightforward to decide for a human, but less quantifiable. A natural choice of method for making a limited-choice decision based on relatively subjective terms is the use of a fuzzy inference system.

2.1. Voxelization pipeline

RT structures are stored in series of 2D planar contours, which can be processed in different ways so that they properly enclose the contoured volume for voxelization. To achieve the highest similarity between the contours and the voxelized volume, we chose to perform the intermediate step of converting the planar contours to a closed surface (Sunderland *et al.* 2015). This step ensures out-of-plane interpolation, handles structures with branches and holes, and applies end-capping as described by Nelms *et al.* (2015). The only input of the triangulation algorithm is the contour set.

The closed surface created using the first step is then used for the voxelization, during which each voxel in the oversampled grid is evaluated whether they are inside or outside the surface. This step has a reference image geometry parameter that is set to be the dose grid, and an oversampling factor parameter, which is determined by the proposed fuzzy-based method.

2.2. Relative structure size

Relative structure size (RSS) is the primary metric based on which the optimal oversampling factor is calculated. We chose to define it as the cubic root of the ratio of the structure volume and the volume of one voxel in the reference dose image, *i.e.* the number of voxels the structure spans on average along each axis.

$$RSS = \sqrt[3]{\frac{structureVolume}{voxelVolume}} \tag{1}$$

The structure volume in Equation 1 is the total volume of the structure in cubic mm, and can be calculated from the closed surface intermediate representation. Voxel volume is the volume of a single voxel, which can be determined as the product of the spacing values along the three axes. Four fuzzy sets describing the relative structure size are shown in Figure 2:

i) Very small – optic lens, optic chiasm, optic nerve, hotspot isodose volume, etc.
ii) Small – brain stem, eyeball, spinal canal, etc.
iii) Medium – bladder, heart, brain, group of small organs, etc.

iv) Large – body, lung, group of medium-sized organs, large isodose volume, etc.

2.3. Structure complexity (compactness)

The topic of quantifying complexity of spatial objects has been studied by several groups. Brinkhoff *et al.* (1995) proposed a 2D contour complexity calculation method mainly for geographical formations. However, the stack of contours representation is a legacy format used in DICOM RT (Mildenberger *et al.*, 2002), and modern formats such as DICOM SEG or research formats are represented as volumes or surface meshes. Furthermore, complexity also needs to be measured in the out-of-plane direction. O'Rourke *et al.* (1985) developed a method for finding minimal enclosing boxes for shapes, which could potentially be useful to apply for complexity measure computation, but aside from the same problem mentioned above, it would also be computationally too complex. Other approaches exist, such as the method of Rigau *et al.* (2005), which uses mutual information to measure object complexity, but it is also too computationally demanding for a task where we actually strive to save computation time by calculating these measures.

Another possibility for quantifying complexity would be to compute a similarity metric – Dice similarity (Dice 1945) or Hausdorff distance (Huttenlocher *et al.* 1993) – of the structure and the sphere centered at the structure's center of mass, with a volume equal to that of the structure. However, these similarity measures require voxelized inputs, so using them would be a major overhead.

The measure of complexity we chose for our algorithm is derived from the work of Alyassin *et al.* (1994), which is specifically designed for medical data, and is implemented in the VTK toolkit (Schroeder *et al.* 2006) in the filter called *vtkMassProperties*. This algorithm calculates the normalized shape index (NSI, see Equation 2 and Figure 5), which characterizes the deviation of the shape of an object from a sphere. NSI is calculated from the surface area and the volume of the structure as

$$NSI = \frac{\sqrt{surfaceArea}}{\sqrt[3]{surfaceVolume}} \left(\frac{\sqrt[3]{4\Pi/3}}{\sqrt{2\Pi}} \right) \tag{2}$$

A sphere's NSI is one by definition due to the normalization expression in Equation 2. The NSI is always greater than or equal to one for complex structures. In experiments conducted on typical anatomical structure sets from the SlicerRT data repository (Pinter *et al.* 2012) (which includes the datasets detailed below) the NSI values fell between 1 and 2. So that the minimum of the complexity measure lies at 0 (instead of 1 which is the minimum of NSI), we define complexity as

$$complexity = NSI - 1 \tag{3}$$

For example, complexity of the eyeballs is zero, and of branching vessels it is 0.6. Two fuzzy sets were defined to describe the complexity of a structure: Low and High (see Figure 2). The fuzzy set parameters were determined based on metrics calculated on a set of structures that were subjectively classified.

2.4. Output oversampling factor

On one hand, too low oversampling factors make an overly crude voxelization grid resulting in unacceptable data loss even for the largest and simplest structures, while the computational gain is less significant with every decrease of the factor. On the other hand if a too large factor value is chosen, the voxelization step will take too long to be feasible, while the geometric gain will become negligible. Thus four possible values were chosen for oversampling factors: ½, 1, 2, and 4. The fuzzy sets of the input measures and the output oversampling factor are reviewed in Figure 2. The input membership functions represent the degree of membership of each structure's quantified property in the defined fuzzy sets.

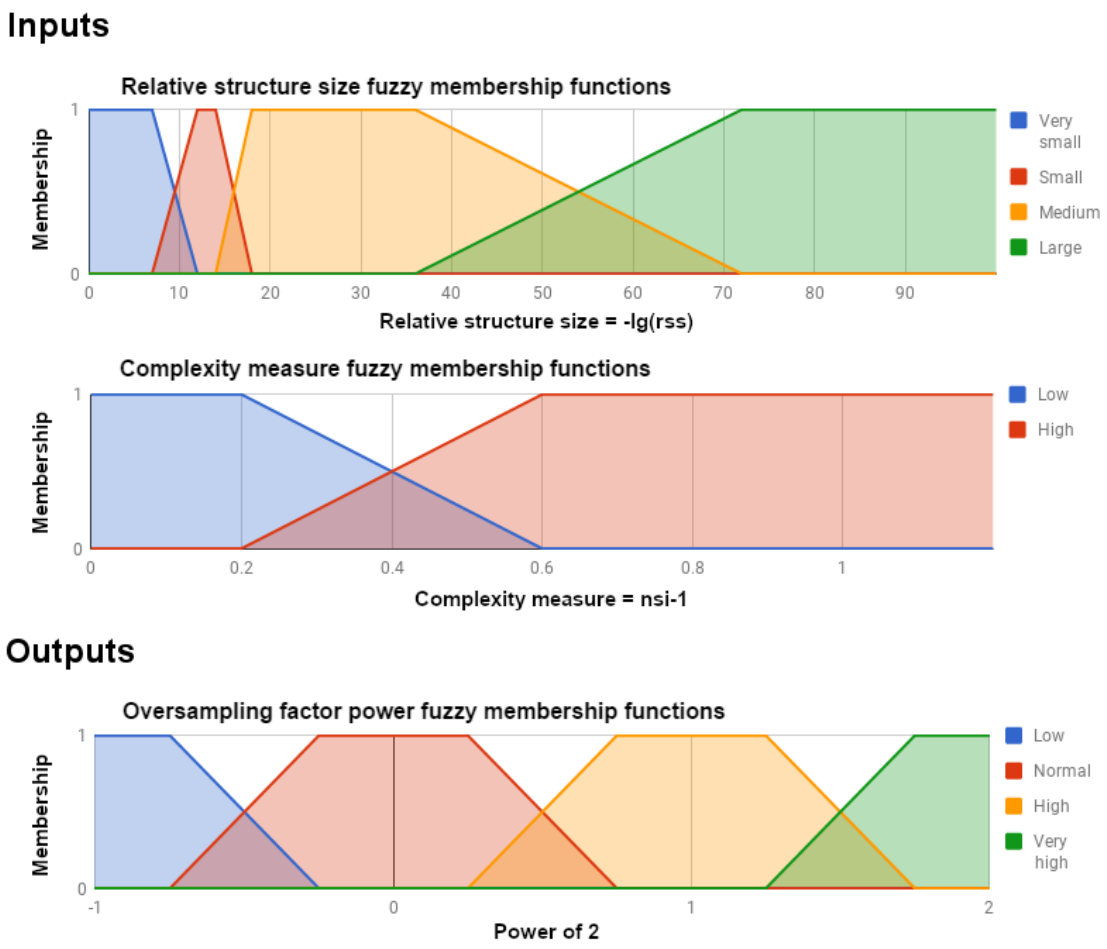

Fig. 2 Fuzzy memberships for the input measures and for the output oversampling factor used in the voxelization process

The membership functions have been established based on the measured metrics (RSS and complexity) and the subjective size and complexity assignments made by the expert clinicians determining the baselines. The system can be considered robust overall against tuning of the parameters. One exception is the separation between the Very small and Small structure size classes, moving which by a value of 1 may mean the difference between an oversampling factor of 2 or 4 for some structures. Thus this value was chosen very carefully considering the structures in the baseline datasets.

2.5. Fuzzy inference system

The two most widely used fuzzy inference systems (Lee 1990) are the Mamdani Model and the Sugeno Model. The primary difference between the two models is in the method for determining a final output value. In a Mamdani Inference System the output of each inference rule is a fuzzy set. To obtain a single output value the output sets are combined and a defuzzification operator is applied. In a Sugeno Inference System the output of each inference rule is a crisp scalar, typically computed as a linear combination of the input variables. The final output value is a weighted average of the output values of the inference rules, where the weights are the degree of applicability of the rules. The Mamdani Model allows for greater flexibility and easier design and comprehension, while the Sugeno Model offers greater computational efficiency (though determining the proper coefficient values in the linear combinations is a potentially difficult problem). In this paper we choose to use a Mamdani Inference System to take advantage of the ease of designing the fuzzy output sets and the greater flexibility offered by the defuzzification stage of the computation. The number of rules we use is small and the computational demands of defuzzification are acceptable. The system defines the following rules:

i) If RSS is Very small, then Oversampling is Very high
ii) If RSS is Small and Complexity is High then Oversampling is High
iii) If RSS is Medium and Complexity is High then Oversampling is High
iv) If RSS is Small and Complexity is Low then Oversampling is Normal
v) If RSS is Medium and Complexity is Low then Oversampling is Normal
vi) If RSS is Large, then Oversampling is Low

First, the input measurements are fuzzified using the membership functions, i.e. the measured exact value is converted into membership value for each set. Next, the rules are applied on them, yielding the output (consequent) functions that are clipped to the membership values, and finally their center of mass is computed, which is the output “crisp” value. This crisp resulting oversampling factor power is rounded, and the actual oversampling factor is calculated as follows:

$$oversamplingFactor = 2^{round(centerOfMass)} \quad (4)$$

Seeing how the inference system works in practice helps in understanding. We choose to illustrate the process through the example of the *Orbit Lt* structure of the *ENT* dataset as seen in Table 1. The input crisp measurements are 8.16 for RSS, and 0.09 for complexity. The 8.16 crisp RSS value yields the fuzzy membership values 0.78, 0.22, 0, and 0 for the RSS fuzzy sets Very small, Small, Medium, and Large, respectively (as can be derived from the functions in Figure 2), and 1 and 0 for the Low and High complexity sets, respectively. Based on the six rules, the consequent values that are used for clipping the output functions are 0.78 for the first rule, 0 for the second rule (because RSS is Small to the degree of 0.22, but complexity is High to the degree of 0, and the 'and' operator takes the minimum of the operands in the Mamdani system), 0 for the third rule, 0.22 for the fourth rule, and 0 for the last two rules. After applying the consequent values to the output sets based on the rules, the consequent functions are formed as seen in Figure 3. The center of mass of the consequent functions 1.26, and so Equation 4 yields the value of 2, which is the defuzzified output.

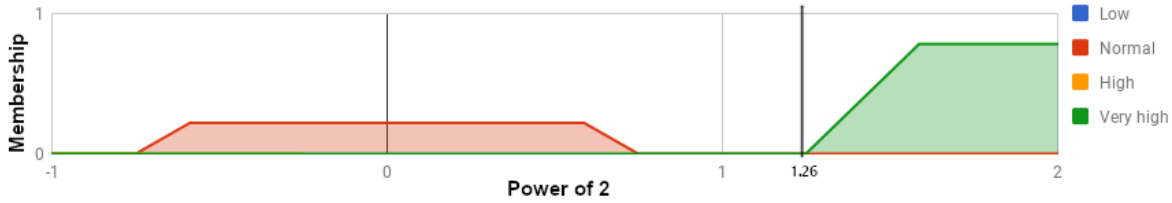

Fig. 3 Consequent functions for the Orbit Lt structure used in the example. The first rule resulted in clipping the Very high output function to 0.78, the fourth rule clipped the Normal function to 0.22, and the rest of the rules resulted in clipping to 0. The center of mass of these consequent functions is 1.26

2.6. Implementation

The fuzzy inference system was initially implemented as part of the SlicerRT toolkit (Pinter *et al.* 2012), which is a general radiation therapy research toolkit. It was later integrated into its host application 3D Slicer (Fedorov *et al.* 2012), an open-source medical image analysis and visualization platform. The fuzzy oversampling calculation algorithm was implemented as one C++ class (*vtkCalculateOversamplingFactor*) that integrates in the general SlicerRT contour handling mechanism and the DVH module.

The only inputs of the algorithm are the intermediate closed surface model and the reference dose image geometry. The steps of the algorithm are illustrated in Figure 4.

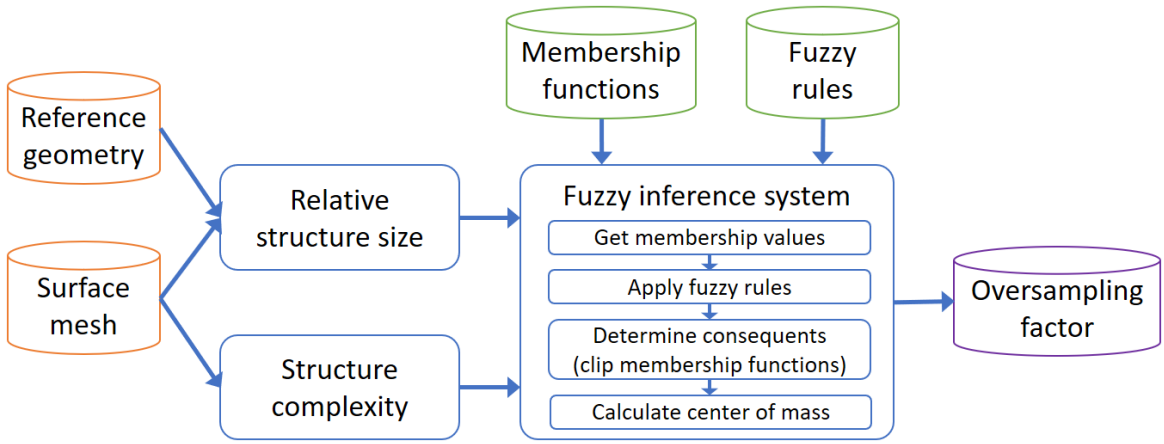

Fig. 4 System diagram of the proposed system. Boxes: algorithms. Orange cylinders (left): input data. Green cylinders (top): parameters of the fuzzy inference system. Purple cylinder (right): output oversampling factor

The system consists of three computational components: relative structure size calculation (see section 2.2), structural complexity calculation (normalized shape index metric, see section 2.3), and the fuzzy inference system discussed in section 2.5. The calculation of the size and complexity metrics involves a single run of the vtkMassProperties algorithm on the input surface mesh. The computational complexity of this execution is *O(n)*, where *n* is the number of triangles in the surface mesh (see Figure 5).

surfaceArea := 0; *MUNC* := 0; *volumeElements* := 0;
for each *triangle* in *surface*:
  get *points* from *triangle*
  get *I*, *J*, and *K* vectors of *triangle*
  calculate *normalVector* of *triangle* from *points*
  increment *MUNC* component based on *normalVector*
  calculate *area* of *triangle* from *I*, *J*, *K* vectors
  increment *surfaceArea* by *area*
  increment *volumeElements* based on *area*, *normalVector*, and *points*
calculate *surfaceVolume* from *MUNC* and *volumeElements*
calculate *NSI* from *surfaceArea* and *surfaceVolume*

Fig. 5 Algorithm calculating the normalized shape index (NSI), determining complexity measure, see Equations 2 and 3. The operations "calculate" and "increment" only use basic operations and square root. Hence the algorithmic complexity is *O(n).*

The algorithm is discussed in detail in the paper of Alyassin *et al.* (1994), and the source code is available in VTK[1]. The outputs of this algorithm are used directly for calculating the structure size and complexity metrics (see Equations 2 and 3), not adding additional computational complexity. The execution of fuzzy inference system mainly involves operations on linear piecewise functions (*i.e.* functions consisting of linear segments). The cost of these operations is *O(n)*, where *n* is the number of segments. This number in the current implementation ranges from 2 to 4, so its computational complexity is negligible. The overall

[1] https://github.com/Kitware/VTK/blob/master/Filters/Core/vtkMassProperties.cxx

complexity of the system is thus *O(n)*, as a function of the number of triangles in the input structure surface mesh.

3. Experimental Results and Discussion

Four RT studies were used to test the method: a treatment plan study for the RANDO® prostate phantom (referred to as "PROS"), a study for the RANDO® ear-nose-throat phantom ("ENT"), a treatment plan for a real patient that contains elaborate contours ("Branching"), and another real patient plan for Fractionated Stereotactic Radiation Therapy (FSRT) treatment that has small targets and high dose gradients ("FSRT"). The results are shown in Table 1.

Structure	Approximate shape	RSS	Subjective size	Complexity measure	Subjective complexity	Oversampling factor	
PROS							
Bladder	spherical	19.59	Med	0.04	Low	1	1
Rectum	cylindrical-vertical	12.53	Med	0.27	Med	2	1 *
Body	cylindrical	88.14	Large	0.09	Low	0.5	0.5
Femoral Head Lt	spherical	14.68	Med	0.03	Low	1	1
ENT							
Brain	spherical	41.47	Med/Large	0.15	Low	1	1
Brain Stem	cylindrical-vert. (banana)	12.63	Small	0.3	Med	2	1 *
Lens Lt	hemisphere	2.06	Very small	0.02	Low	4	4
Optic Chiasm	"bone-shaped"	4.19	Very small	0.05	Med	4	4
Optic Nerve Lt	cylindrical-horizontal	4.82	Very small	0	Low/Med	4	4
Orbit Lt	spherical	8.16	VeryS/Small	0.09	Low	2	2
PTV1	spherical	20.10	Med	0.08	Low	1	1
Branching							
Body	cylindrical	145.70	Large	0.38	Med	0.5	0.5
Kidney Rt	kidney	25.76	Med	0.23	Low/Med	2	1 *
Vessels	branching cylindrical (Y)	27.13	Med	0.99	High	2	2
Spinal Canal	cylindrical-vertical-long	15.65	Small/Med	0.52	Med/High	2	2
Liver	liver	48.50	Med/Large	0.47	Med/High	1	1
FSRT							
Body	cylindrical	92.49	Large	0.15	Low	0.5	0.5
Brain	spherical	59.78	Med/Large	0.08	Low	1	1
Brain Stem	cylindrical	15.96	Small	0.2	Med	1	1
Cochlea Lt	cylindrical	6.03	Very small	0.06	Low	4	4
GTV1 32.5/5	spherical	7.26	VeryS/Small	0.11	Low	4	4
GTV2 30.0/5	spherical	4.51	Very small	0.08	Low	4	4
Lens Lt	hemisphere	3.34	Very small	0.16	Low	4	4
Optic Chiasm	bone-shaped	4.26	Very small	0.34	Med	4	4

Optic Nerve Lt	cylindrical-horizontal	5.14	Very small	0.31	Low/Med	4	4
Orbit Lt	spherical	10.50	VeryS/Small	0.02	Low	2	2
PTV1	spherical	10.17	VeryS/Small	0.06	Low	2	2
PTV2	spherical	7.28	VeryS/Small	0.04	Low	4	4
Skin	thin layer	51.08	Med/Large	1.35	High	1	1
Spinal Canal	cylindrical-vert. (banana)	14.27	Small	0.31	Med	1	1

Table 1. Test structures, their computed and subjectively determined measures, and their computed oversampling factors. Oversampling factor color-codes: blue – baseline, green – perfect match, yellow (*) – not optimal but acceptable.

Baseline oversampling factors for each test structure were determined as consensus by several clinicians in the cancer program at the Kingston Health Sciences Centre. Each of the participating clinicians can be considered an expert in RT treatment planning, and the decision was made based on their personal experience. The decision regarding the optimal oversampling factor involved considering the following properties of the individual structures: typical size and complexity, variance in size and shape within the patient population, and proximity of the structure to high dose gradients. The calculated oversampling factors using the three different surface representations were compared to the consensus clinical decision. The results show good matching overall, with most of the cases matching perfectly, few acceptable deviations, and no failures.

Oversampling factor calculation was also tested on the synthetic test dataset used by Nelms et al. (2015). The four dose volumes in the dataset had the resolutions of 0.4x0.4x0.2mm^3, isotropic 1mm, 2mm, and 3mm. The 50 structures were converted to closed surface as described in 2.1. The system calculated oversampling factors for all dose-structure pairs. The results are shown in Table 2.

Dose resolution (mm)	**Number of structures with oversampling factor**			
	0.5	**1**	**2**	**4**
0.4 x 0.4 x 0.2	20	30	-	-
1 x 1 x 1	-	50	-	-
2 x 2 x 2	-	20	30	-
3 x 3 x 3	-	-	6	44

Table 2. Oversampling factors calculated on the Nelms *et al.* 2015 dataset. Each cell contains the number of structures being assigned an oversampling factor paired with the different dose volumes.

Unsurprisingly, the resulting oversampling factors were mostly influenced by the resolution of the reference dose volume. In case of the dose volumes where one oversampling factor was assigned to some structures and another to others, the lower oversampling factor was calculated without exception for

cylinder structures created on low resolution dose volumes. One reason for this divide is that these structures were considered low complexity due to end-capping, which added a top and bottom circular "slice" with smaller diameter. The other reason is that the cylinders contain a larger number of voxels than the spheres and cones, and these two factors contributed to lower outputs. We can say that the results obtained on the validation datasets show oversampling values as expected.

The DVHs computed using automatic oversampling factors and a constant manual factor of two were also compared to results of third party applications via SlicerRT's DVH comparison module. The comparison algorithm follows the gamma-like DVH comparison method described by Ebert *et al.* (2010). We used the results from two applications, one being Computational Environment for Radiotherapy Research (CERR) (O'Deasy *et al.* 2003), a popular Matlab-based tool, and Eclipse® (Varian Medical, Palo Alto, CA, USA), a commercial RT treatment planning system that is used in many hospitals and clinics all over the world. According to the CERR documentation no oversampling is performed, which means the software uses a factor of 1. Unfortunately we were not able to access such documentation for Eclipse®, and the software gives no information about the DVH calculation details. However, its DVH calculation has been validated (Gossman *et al.* 2010), and was found to satisfy the requirements set by the authors, so can be considered a suitable baseline for this study. The comparison results (see Table 3) show that the new method utilizing automatic oversampling factor calculation provides a similar match to the DVHs calculated by the third-party applications than the use of a fixed oversampling factor of two. In most comparisons the automatic oversampling factor performs better than the fixed factor of two, however, general improvement cannot be concluded. The goal of the comparison was to provide a wide range of accuracy values for better comparison, which was achieved by applying very strict tolerance values to be able to detect even the smallest deviations (1%/1mm instead of the typical gamma tolerances 3%/3mm).

Structure	Eclipse® w/ Auto	Eclipse® w/ Manual	CERR w/ Auto	CERR w/ Manual
PROS Overall	95.74	97.53	99.93	97.68
BODY	100	100	100	100
Bladder	100	100	100	100
Femoral Head Lt	99.90	99.90	99.58	100
Femoral Head Rt	98.14	99.90	100	100
PTV	99.61	99.61	100	100
Rectum	76.76	85.74	100	86.08
ENT Overall	95.64	95.23	95.60	95.82
BODY	100	100	100	100
BRAIN	100	100	100	100
BRSTEM	100	99.04	93.31	100
CTV	100	100	100	100
GTV	98.95	98.95	99.04	99.04
Lens Lt	98.56	98.56	98.73	98.73
Lens Rt	99.52	99.52	99.36	99.68
Optic Chiasm	87.74	82.18	85.67	82.48
Optic Nerve Rt	86.49	86.59	85.99	85.99
Optic Nerve Lt	85.82	85.73	85.67	85.67
Orbit Lt	98.08	98.08	99.73	98.73
Orbit Rt	98.95	98.95	98.73	98.73
PTV	100	100	100	100
optBRAIN	100	100	100	100
optOptic	90.42	90.42	98.41	98.41
FSRT Overall	97.14	94.24	95.65	95.07
Body	100	100	100	100
Brain	100	100	100	100
Brain Stem	100	100	100	100
Cochlea Lt	99.17	99.14	98.97	98.97
Cochlea Rt	99.53	99.51	99.48	99.48
GTV1	100	100	93.81	92.78
GTV2	95.87	96.28	93.30	93.30
Lens Lt	93.19	93.79	92.27	92.27
Lens Rt	96.28	96.75	95.36	95.36
Optic Chiasm	94.36	94.54	90.72	88.66
Optic Nerve Lt	78.15	79.42	79.90	73.20
Optic Nerve Rt	98.18	98.42	92.78	92.78
Orbit Lt	100	100	99.48	99.48
Orbit Rt	100	100	97.42	97.42
PTV1	97.14	97.14	97.94	97.94
PTV2	96.86	95.53	90.21	89.69
Skin	100	100	100	100
Spinal Canal	99.84	99.87	100	100

Table 3. DVH comparison for SlicerRT manual oversampling of 2 (columns with Manual) and automatic oversampling (columns with Auto), to CERR and Eclipse®

The computation time of the oversampling calculation algorithm itself was also assessed, by measuring the time needed to perform the two input measurements, and the time needed to run the fuzzy inference system. 178 measurements were made with all the three types of input surfaces and all the test structures. The summarized average computation times can be seen in Table 4.

	Total time (ms)	Measures (ms)	Fuzzy rules (ms)
Mean	16.5	5.9	10.6
Std.Dev.	16.1	14.5	4.1
Min	6.0	0.0	5.0
Max	112.0	99.0	37.0

Table 4. Overall averaged time measurements (in milliseconds) for calculating the automatic oversampling factor on each structure listed in Table 1.

Comparison of the computation times in SlicerRT of the DVHs with automatic oversampling and manual oversampling of two was also performed. The measured times include calculation of the oversampling factor (in case of automatic), voxelization using the oversampling factor, and calculation of the DVH itself. Each calculation was run five times. The mean and standard deviation, as well as the improvement is summarized in Table 5. It can be seen that the computation times improved in every case, and in case of the datasets with limited number of structures falling in the very small category, the time needed to do the entire calculation was halved.

	Auto (s)		Manual (s)		Improvement
	Mean	Std.Dev.	Mean	Std.Dev.	in mean (%)
PROS	0.93	0.04	1.56	0.02	0.40
ENT	1.40	0.01	2.74	0.07	0.49
Branching	5.52	0.03	11.90	0.29	0.54
FSRT	16.53	0.18	22.46	1.00	0.26

Table 5. Comparison of calculating DVH in SlicerRT using automatic and manual oversampling. The measured times include: calculation of the oversampling factor (only for automatic); voxelization using the oversampling factor; calculating the DVH

3D Slicer can be freely downloaded[2] for Windows, Mac, and Linux 64-bit operating systems. Although the inference system and the voxelization algorithms are included in 3D Slicer core, the SlicerRT extension is needed to be able to use it for DVH calculation. It can be installed from the Extension Manager in 3D Slicer. SlicerRT versions 0.15.2 and above contain the described method. Information about usage can be found on the SlicerRT website[3]. The software comes with BSD-style license, so it is freely usable, modifiable and distributable.

3.1. Discussion

We recognize that the most robust validation setting would be against the corresponding components of commercial treatment planning systems, such as Eclipse® or Pinnacle³ (Philips, Koninklijke Philips N.V, Netherlands). However, without disclosed implementation details or source code, the atomic algorithms

[2] http://download.slicer.org
[3] http://slicerrt.org

are not accessible for comparative execution. Thus it is unfeasible to assess DVH quality without the knowledge of the properties of the structure volume data, which is the direct input of the DVH calculation. Similarly, it is also impossible to compare computation speeds, as although thousands of DVHs are calculated in each plan optimization process, DVH calculation happens among many other operations, without the possibility to isolate the DVH calculation and related operations. The proprietary nature of these software also prevents showing any improvements beyond their capabilities in the comparative validation. The DVHs calculated using the proposed method for very small and/or high complexity structures, which are most prone to data loss during voxelization are potentially more accurate than the baselines used for validation, so any improvement manifests itself as discrepancy to these baselines.

A possible improvement could be using other complexity measures. The current one only calculates a global complexity measure, but we may also be interested in local complexity, so that locally highly complex regions (“zig-zags”) are not oversimplified during voxelization.

It would be also interesting to see how odd oversampling values perform versus the currently used even numbers. As the paper by Nelms *et al.* (2015) suggests that using odd values ensures that “no slabs of ‘new’ voxels (that will need to be interpolated) are centered exactly between the original dose plane locations”, this consideration might further improve the accuracy of the voxelized structure volumes. For accuracy measurement, a vertex-based Hausdorff evaluation would be also useful in addition to comparative DVH validation.

4. Conclusion

Accurate and fast DVH calculation is essential for dose evaluation in both clinical and research settings. It is even more so if inverse treatment planning is performed based on thousands of calculated DVHs, where small or complex structures involved in the dose-volume objectives may skew the optimization process due to voxelization inaccuracies, potentially resulting in sub-optimal treatment plans.

A fuzzy-based method was implemented for automatically determining oversampling factor for each processed structure to facilitate accurate DVH calculation while keeping computational cost reasonable. The results show that the proposed algorithm reliably identifies the small and complex structures on various input types and on a wide range of clinical structures including the extreme cases. Considerable savings of computation times can be achieved by using the oversampling factor automatically determined by the fuzzy inference system, while keeping the DVH accuracy very similar, compared to the assessed third party applications.

Widening the range of possible output oversampling factors in the future may show further improvements in accuracy. Also, including analytical datasets besides clinical datasets in DVH validation would increase confidence in the accuracy of the system.

In conclusion, the proposed method potentially facilitates more dependable dose-volume based optimization methods, while yielding a good balance between metric accuracy and calculation time, both critical in the clinic.

5. Acknowledgement

The authors thank Dr. Robin Dawes for providing insight in the design and justification of the fuzzy inference system.

This is a post-peer-review, pre-copyedit version of an article published in Soft Computing. The final authenticated version is available online at https://doi.org/10.1007/s00500-020-05126-w.

6. Compliance with Ethical Standards

Funding: This research was funded in part by the Ontario Consortium for Adaptive Interventions in Radiation Oncology (OCAIRO), and the Applied Cancer Research Unit program of Cancer Care Ontario with funds provided by the Ontario Ministry of Health and Long-Term Care. Gabor Fichtinger was supported as a Cancer Care Ontario Research Chair in Cancer Imaging.

Conflict of interest: All authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Alyassin, A. M., Lancaster, J. L., Downs, J. H., & Fox, P. T. (1994). Evaluation of new algorithms for the interactive measurement of surface area and volume. Medical Physics, 21(6), 741-752.

Babier, A., Boutilier, J. J., Sharpe, M. B., McNiven, A. L., & Chan, T. C. (2018). Inverse optimization of objective function weights for treatment planning using clinical dose-volume histograms. Physics in Medicine & Biology, 63(10), 105004.

Brinkhoff, T., Kriegel, H. P., Schneider, R., & Braun, A. (1995, December). Measuring the Complexity of Polygonal Objects. In ACM-GIS (p. 109).

Corbett, J. F., Jezioranski, J., Crook, J., & Yeung, I. (2002). The effect of voxel size on the accuracy of dose-volume histograms of prostate 125I seed implants. Medical Physics, 29(6), 1003-1006.

Das, I. J., Cheng, C. W., Chopra, K. L., Mitra, R. K., Srivastava, S. P., & Glatstein, E. (2008). Intensity-modulated radiation therapy dose prescription, recording, and delivery: patterns of variability among institutions and treatment planning systems. Journal of the National Cancer Institute, 100(5), 300-307.

Deasy, J. O., Blanco, A. I., & Clark, V. H. (2003). CERR: a computational environment for radiotherapy research. Medical Physics, 30(5), 979-985.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. Ecology, 26(3), 297-302.

Drzymala, R. E., Mohan, R., Brewster, L., Chu, J., Goitein, M., Harms, W., & Urie, M. (1991). Dose-volume histograms. International Journal of Radiation Oncology* Biology* Physics, 21(1), 71-78.

Ebert, M. A., Haworth, A., Kearvell, R., Hooton, B., Hug, B., Spry, N. A., ... & Joseph, D. J. (2010). Comparison of DVH data from multiple radiotherapy treatment planning systems. Physics in Medicine and Biology, 55(11), N337.

Eisemann, E., & Décoret, X. (2008). Single-pass GPU solid voxelization for real-time applications. In Proceedings of graphics interface 2008 (pp. 73–80).

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M. and Buatti, J., 2012. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imaging, 30(9), pp.1323-1341.

Fraass, B., Doppke, K., Hunt, M., Kutcher, G., Starkschall, G., Stern, R., & Van Dyke, J. (1998). American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. Medical Physics, 25(10), 1773-1829.

Gossman, M. S., & Bank, M. I. (2010). Dose-volume histogram quality assurance for linac-based treatment planning systems. Journal of medical physics/Association of Medical Physicists of India, 35(4), 197.

Hazell, I., Bzdusek, K., Kumar, P., Hansen, C. R., Bertelsen, A., Eriksen, J. G., ... & Brink, C. (2016). Automatic planning of head and neck treatment plans. Journal of applied clinical medical physics, 17(1), 272-282.

Huang, J., Yagel, R., Filippov, V., & Kurzion, Y. (2014). An accurate method for voxelizing polygon meshes. IEEE Symposium on Volume Visualization (Cat. No.989EX300), (October), 119-126.

Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(9), 850-863.

J. Leasure, V. Ulizio and D. Pearson. Isodose Volume Comprison of Two Common Volume Derivation Methods. Medical Physics 44(6) 2961 (2017)

Kirisits, C., Pötter, R., Lang, S., Dimopoulos, J., Wachter-Gerstner, N., & Georg, D. (2005). Dose and volume parameters for MRI-based treatment planning in intracavitary brachytherapy for cervical cancer. International Journal of Radiation Oncology* Biology* Physics, 62(3), 901-911.

Kooy, H. M., Nedzi, L. A., Alexander, E., Loeffler, J. S., & Ledoux, R. J. (1993). Dose-volume histogram computations for small intracranial volumes. Medical physics, 20(3), 755-760.

Lee, C. C. (1990). Fuzzy logic in control systems: fuzzy logic controller. I. IEEE Transactions on Systems, Man, and Cybernetics, 20(2), 404-418.

Makishima, H., Ishikawa, H., Terunuma, T., Hashimoto, T., Yamanashi, K., Sekiguchi, T., ... & Sakurai, H. (2015). Comparison of adverse effects of proton and X-ray chemoradiotherapy for esophageal cancer using an adaptive dose–volume histogram analysis. Journal of radiation research, 56(3), 568-576.

Mildenberger, P., Eichelberg, M., Martin, E. (2002). Introduction to the DICOM standard. Eur. Radiol. 12, 920–927.

Mountris, K. A., Visvikis, D., & Bert, J. (2019). DVH-Based Inverse Planning Using Monte Carlo Dosimetry for LDR Prostate Brachytherapy. International Journal of Radiation Oncology* Biology* Physics, 103(2), 503-510.

Nelms, B., Stambaugh, C., Hunt, D., Tonner, B., Zhang, G., & Feygelman, V. (2015). Methods, software and datasets to verify DVH calculations against analytical values: Twenty years late (r). Medical Physics, 42(8), 4435-4448.

Novotný, P., Dimitrov, L. I., & Šrámek, M. (2010). Enhanced voxelization and representation of objects with sharp details in truncated distance fields. IEEE Transactions on Visualization and Computer Graphics, 16(3), 484–498.

O'Rourke, J. (1985). Finding minimal enclosing boxes. International Journal of Parallel Programming, 14(3), 183-199.

Pinter, C., Lasso, A., Wang, A., Jaffray, D., & Fichtinger, G. (2012). SlicerRT: Radiation therapy research toolkit for 3D Slicer. Medical Physics, 39(10), 6332-6338.

Rigau, J., Feixas, M., & Sbert, M. (2005, June). Shape complexity based on mutual information. In Shape Modeling and Applications, 2005 International Conference (pp. 355-360). IEEE.

Rodrigues, G., Lock, M., D'Souza, D., Yu, E., & Van Dyk, J. (2004). Prediction of radiation pneumonitis by dose–volume histogram parameters in lung cancer—a systematic review. Radiotherapy and oncology, 71(2), 127-138.

Rose, J., Rodrigues, G., Yaremko, B., Lock, M., & D'Souza, D. (2009). Systematic review of dose–volume parameters in the prediction of esophagitis in thoracic radiotherapy. Radiotherapy and Oncology, 91(3), 282-287.

Schroeder, W. J., Lorensen, B., & Martin, K. (2004). The visualization toolkit: an object-oriented approach to 3D graphics. Kitware.

Schwarz, M., & Seidel, H.-P. (2010). Fast parallel surface and solid voxelization on GPUs. ACM Transactions on Graphics (TOG), 29(6), 179.

Sunderland, K., Woo, B., Pinter, C., & Fichtinger, G. (2015). Reconstruction of surfaces from planar contours through contour interpolation. In Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 9415). [94151R] SPIE.

Sunderland, K., Pinter, C., Lasso, A., & Fichtinger, G. (2016). Effects of voxelization on dose volume histogram accuracy. In Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 9786). [97862O] SPIE.