# **ORIGINAL RESEARCH**





Comparison of Otsu and an adapted Chan-Vese method to determine thyroid active volume using Monte Carlo generated SPECT images

Jonas Högberg<sup>1</sup>, Christoffer Andersén<sup>2</sup>, Tobias Rydén<sup>3</sup> and Jakob H. Lagerlöf<sup>2,4,5\*</sup>

\*Correspondence: jakob.lagerlof@regionvarmland. se

<sup>1</sup> Department of Medical Radiation Physics, and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

<sup>2</sup> Department of Medical Physics, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

<sup>3</sup> Department of Medical Physics and Biomedical Engineering. Sahlgrenska University Hospital, Gothenburg, Sweden <sup>4</sup> Department of image and Functional Diagnostics, Karlstad Central Hospital, Karlstad, Sweden <sup>5</sup> Centre for clinical research and education, Region Värmland, Karlstad, Sweden

# Abstract

Background: The Otsu method and the Chan–Vese model are two methods proven to perform well in determining volumes of different organs and specific tissue fractions. This study aimed to compare the performance of the two methods regarding segmentation of active thyroid gland volumes, reflecting different clinical settings by varying the parameters: gland size, gland activity concentration, background activity concentration and gland activity concentration heterogeneity.

Methods: A computed tomography was performed on three playdough thyroid phantoms with volumes 20, 35 and 50 ml. The image data were separated into playdough and water based on Hounsfield values. Sixty single photon emission computed tomography (SPECT) projections were simulated by Monte Carlo method with isotope Technetium-99 m (<sup>99m</sup>Tc). Linear combinations of SPECT images were made, generating 12 different combinations of volume and background: each with both homogeneous thyroid activity concentration and three hotspots of different relative activity concentrations (48 SPECT images in total). The relative background levels chosen were 5 %, 10 %, 15 % and 20 % of the phantom activity concentration and the hotspot activities were 100 % (homogeneous case) 150 %, 200 % and 250 %. Poisson noise, (coefficient of variation of 0.8 at a 20 % background level, scattering excluded), was added before reconstruction was done with the Monte Carlo-based SPECT reconstruction algorithm Sahlgrenska Academy reconstruction code (SARec). Two different segmentation algorithms were applied: Otsu's threshold selection method and an adaptation of the Chan–Vese model for active contours without edges; the results were evaluated concerning relative volume, mean absolute error and standard deviation per thyroid volume, as well as dice similarity coefficient.

**Results:** Both methods segment the images well and deviate similarly from the true volumes. They seem to slightly overestimate small volumes and underestimate large ones. Different background levels affect the two methods similarly as well. However, the Chan–Vese model deviates less and paired *t*-testing showed significant difference between distributions of dice similarity coefficients (p-value < 0.01).



© The Author(s) 2024. 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/.

**Conclusions:** The investigations indicate that the Chan–Vese model performs better and is slightly more robust, while being more challenging to implement and use clinically. There is a trade-off between performance and user-friendliness.

**Keywords:** Image segmentation, Monte Carlo, SPECT, Thyroid volume, Radioiodine therapy, Otsu, Chan–Vese

# Background

The thyroid gland, which is shaped like a butterfly, is located anterior to the trachea and plays a critical role in regulating the body's metabolism through the production of the hormones Triiodothyronine (T3) and Thyroxine (T4). The volume of the thyroid gland can vary due to natural causes, but it is also influenced by various factors such as gender, age, height, weight, iodine intake, smoking, and environmental circumstances. The relationship between the thyroid gland volume and these characteristics is known to be nonlinear. In a healthy adult without iodine deficiency, the mean sonographic volume of the thyroid gland is estimated to be 7 to 10 ml [1].

The production of T3 and T4 hormones in the thyroid gland depends on the oxidation and binding of iodine to tyrosine. T4, the hormone produced exclusively in the thyroid gland, can be considered a prohormone to T3, as the latter, being the more potent and metabolically active hormone, can be synthesised from T4. Some T3 is secreted directly into the bloodstream, but the largest amount of hormone secreted from the thyroid gland is in the form of T4. The synthesis of T4 to T3 is then carried out by the liver, pituitary gland, and other endocrine organs. Enlargement of the thyroid gland, hyperthyroidism (also known as a goitre), can be caused by various factors including autoimmune diseases (such as Graves' disease), toxic nodular goitre, cancer, hereditary factors, medication, and diets lacking sufficient amounts of iodine [2].

Hyperthyroidism manifests itself physically as a moderate to severe increase in thyroid volume; a large goitre - 80 g (80 ml) and above - is a contraindication for radioiodine therapy [3, 4]. Thyroid volumes of 40–60 ml may also be considered for surgical procedures, depending on other clinically relevant parameters [5, 6]. Radioiodine treatment has been shown to be effective even within this volume range [7]. Patients with small goitres, which are slightly larger than a normal-sized thyroid gland, are most often treated with antithyroid drugs (ATD), but may later be considered for radioiodine therapy, if 12–18 months of ATD treatment proves ineffective [3, 5, 8].

Different methods are available for assessing the volume of the thyroid gland, including palpation (physical examination), 2D and 3D ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and planar scintigraphy (PS). A study by Viduetsky et al. [9] shows that MRI has the best precision with an error of less than 4 %. This volume measurement is useful for surgical planning and monitoring the effectiveness of medication changes. For the individualised planning of radioactive treatment of thyrotoxicosis by Iodine-131 (<sup>131</sup>I) (radioiodine therapy), single-photon emission computed tomography (SPECT) can be utilised as the imaging modality to determine the active volume of the thyroid gland. SPECT imaging is preferably performed with Technetium-99 m (<sup>99m</sup>Tc) pertechnetate: the uptake levels of <sup>99m</sup>Tc pertechnetate and <sup>131</sup>I are highly correlated, whereas <sup>99m</sup>Tc SPECT imaging offers better quality, due to lower gamma photon energies, resulting in a lower patient absorbed radiation dose than from <sup>131</sup>I SPECT imaging [10]. Despite the critical importance of accurate volume determination for the effective prescription of radioactive iodine, there is a lack of consensus among hospitals in Sweden on the preferred method for calculating thyroid volume and <sup>99m</sup>Tc uptake levels from SPECT or PS images [11]. For the planning of radioiodine therapy of the thyroid, both functional volume, as well as pertechnetate uptake need to be quantified, and SPECT is considered the best way of achieving this [12–15] - neither can 3D shape of active volume be easily described by PS, nor can activity inside the thyroid be distinguished from activity in front of or behind the gland. Thus, thyroid pertechnetate uptake is often overestimated by PS imaging [11, 16].

SPECT images are challenging to segment compared to other tomographic images, such as CT or MRI, due to their low resolution and inherent blurry characteristics. On the other hand, SPECT provides valuable physiological information that is not available from CT or MRI.

A common segmentation approach when SPECT is used is intensity thresholding of the images. This is an arbitrary approach as every patient case is unique, while the threshold values are either fixed or operator-dependent. Apart from being accurate, robust, and operator-independent, a segmentation method for clinical implementation should be easy to use [17].

Two methods that have been successful in determining the volume of other organs are the Otsu method [18] and the Chan–Vese (C–V) model [19–21]. As the volume of the thyroid gland plays a crucial role in radioiodine treatment planning, it is important to assess the accuracy of these methods for segmenting the thyroid gland. The aim of this study is to compare the performance of the Otsu method and the C–V model in calculating the active volume of thyroid glands, taking into account variations in size, activity concentration (AC), heterogeneity, and background AC. The study seeks to determine which method is more accurate and robust in calculating the active volume of thyroid glands, and therefore most suitable for clinical implementation: the Otsu method or the Chan–Vese model.

As radioiodine therapy is considered mainly for moderately sized goitres, typically with a total thyroid gland volume of 20-40 ml and, only in rare cases, 60 - 80 ml, this study focuses on the most relevant thyroid volume range, i.e. 20 - 50 ml [3-6].

# Methods

#### Image generation

To generate images, a computed tomography (CT) was performed of the three playdough thyroid phantoms described in [22], with volumes 20, 35 and 50 ml, using a pixel size of 0.98 by 0.98 mm and slice thickness of 2.5 mm. The formula for the playdough was 38% common salt (NaCl), 36% water, 21% wheat flour and 5% rapeseed oil [22]. The thyroid phantoms were fixed on an air-filled tube (trachea), wrapped in a plastic film, and immersed in water to simulate surrounding tissues (see Fig. 1). The CT images were resampled to 4.42 mm cubic voxels and annotated playdough, water and air, based on thresholding of hounsfield values (600, 0 and -1000, respectively). The threshold values



Fig. 1 Playdough phantom A playdough fantom fixed on an air-filled plastic tube, mounted inside a water tank. The water level in the tank will reach just above the playdough

were set to match the true volumes as accurately as possible. The volumes of the annotated playdough phantoms were 20.03, 34.97 and 50.00 ml.

Sixty single photon emission computed tomography (SPECT) projections were generated using the Monte Carlo (MC) method with the <sup>99m</sup>Tc isotope [23]. The MC algorithm is GPU-based and uses delta scattering, forced interaction, scattering orders (0 to 2) and forced detection for variance reduction. A pre-generated angular response function (ARF), for all angles and energies, was used to model collimator response. For each phantom, primary and scattered radiation were simulated separately for both the background and phantom. The different segments (water, thyroid and air) were assigned attenuation values. The same values as for water were also used for the thyroid voxels. Linear combinations of these SPECT images were created to generate 12 different combinations of volumes and backgrounds, with homogeneous thyroid AC and a hotspot of three different relative ACs, resulting in a total of 48 SPECT images.

The relative background levels were set at 5 %, 10 %, 15 %, and 20 % of the phantom AC, while the hotspot activities were set at 100 % (homogeneous case), 150 %, 200 %, and 250 %. Because variance reduction was used in the simulations, Poisson noise, sampled from a Poisson distribution based on the pixel value, was added before reconstruction [24], resulting in a coefficient of variation of 0.8 in the image background of the reconstructed image, in the case of the 20 % background (with the scattering excluded).

A Low energy high-resolution (LEHR) collimator was used, with thickness 35 mm, hole diameter 1,5 mm, septal thickness 0.2 mm, crystal thickness 15.875 mm and intrinsic resolution, full width at half maximum (FWHM), 4.5 mm. The maximum thyroid-detector distance was 100 mm and the system resolution (FWHM) was calculated, according to [25], to be 7.5 mm.

An average of 100,000 counts were recorded in each projection. The matrix size was 128 x 128, with 4.42 mm cubic voxels. The MC-based SPECT reconstruction algorithm Sahlgrenska Academy reconstruction code (SARec) [23] was used for reconstruction, with six subsets and twelve iterations, and no post-filters were applied. SARec is a GPU-based maximum-likelihood expectation-maximisation (MLEM) algorithm that uses the above MC-algorithm to simulate SPECT projections from an estimated activity distribution. For each iteration, the activity estimates are updated through back-projection of the ratio of simulated to measured projection (which in this study is MC-simulated).

# Image pre-processing

For efficient segmentation, it is desirable to have a homogeneous and low background. To achieve this, a background region of interest (ROI) was selected and its mean value was subtracted from the entire image to equate background tissue AC to the surrounding air, which is not radioactive. To prevent potential hotspots from being segmented separately and misclassifying the remaining thyroid tissue as background, a representative ROI of thyroid tissue was outlined, and its mean value was used as an upper limit for values in the image. Although wavelet-based denoising [26] was initially applied, it was found to be irrelevant to the segmentation result and was therefore omitted.

# Image segmentation

The images were imported into MATLAB [27] as 8-bit unsigned integer matrices (values between 0 and 255). Two different segmentation algorithms were applied: 1) Otsu's threshold selection method and 2) the iterative convolution-thresholding method (ICTM) for active contours without edges, developed from the C–V model, by Wang et al. [28]. The results were evaluated based on the relative volume, mean absolute error (MAE), calculated using Eq. 1, standard deviation (SD) (Eq. 2), and dice similarity coefficient (DSC) of the segmented objects (Eq. 3) per thyroid volume. The DSC measures the similarity between the calculated segmentation and the ground truth segmentation, with a score of 1 indicating perfect agreement and a score of 0 indicating no agreement. A paired *t*-test was conducted to see if there was a significant difference ( $\alpha = 0.05$ ) between the DSC for the two methods. To investigate the impact of background on segmentation performance, DSC was calculated for background levels between 0 % and 70 % for a 50 ml phantom without hotspot.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |\hat{V}_i - V_i|$$
(1)

 $\hat{V}_i$  and  $V_i$  are the estimated and true volumes, respectively, of each of the N objects.

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\hat{V}_i - \overline{V})^2}$$
(2)

S is the sample standard deviation and  $\overline{V}$  is the mean relative volume (MRV).

$$DSC = \frac{2|V \cap \hat{V}|}{|V| + |\hat{V}|} \tag{3}$$

*V* is the volume from the CT images and  $\hat{V}$  are the estimates by Otsu or Chan–Vese.

# Otsu's method for image thresholding

Otsu's method is a well-established segmentation algorithm that was developed in the 1970s [18], and is still in use today. This global thresholding algorithm selects the optimum threshold by minimising the weighted within-class variance (or maximising the between-class variance) of the thresholded background and foreground voxels in the image (Eq. 4).

$$\sigma_w^2(t) = \omega_b(t)\sigma_b^2(t) + \omega_f(t)\sigma_f^2(t) \tag{4}$$

 $\omega_b$  and  $\omega_f$  are the weights of the background and foreground classes, separated by a threshold t, and  $\sigma_b^2$  and  $\sigma_f^2$  are variances of the classes. The class weight  $\omega_{b,f}(t)$  is computed from the 256 bins of the histogram (Eq. 5).

$$\omega_b(t) = \sum_{i=0}^{t-1} p(i)$$

$$\omega_f(t) = \sum_{i=t}^{L-1} p(i)$$
(5)

*L* is the total number of greyscale levels and p(i) are the normalised counts of each bin (greyscale level). The threshold level (*t*) corresponding to the minimum  $\sigma_w^2(t)$  was selected.

# Chan-Vese model for active contours

The C–V model for active contours without edges is an effective segmentation algorithm suitable for objects without distinct borders, as it does not rely on gradients in image intensity [29]. Additionally, it is tolerant to blurring and noise [29], making it appropriate for medical imaging applications. The algorithm is based on the Mumford and Shah functional for image segmentation, which was first introduced in 1985 [30–32].

This paper will not provide a complete mathematical explanation of the C–V model, but the principle of it is to minimise the weighted sum of the total sums of squares for the object and the background, which corresponds to the total voxel-by-voxel difference from the mean within the object and background, respectively. The weights that are set by the user for the algorithm to converge depend on the heterogeneity of the current image, and the user needs to define an arbitrary initial contour. To encourage a smoother surface, regularisation terms are added to penalise large object volume and surface area. The energy functional to minimise, denoted as  $F(c_1, c_2, C)$  (Eq. 6),

$$F(c_1, c_2, C) = \mu \cdot area(C) + \nu \cdot volume(inside(C)) + \lambda_1 \int_{inside(C)} |u_0(x, y, z) - c_1|^2 dx dy dz + \lambda_2 \int_{outside(C)} |u_0(x, y, z) - c_2|^2 dx dy dz,$$
(6)

where *C* is the surface enclosing the object,  $\mu$ ,  $\nu$ ,  $\lambda_1$  and  $\lambda_2$  are constants,  $u_0(x, y, z)$  are individual voxel values,  $c_1$  and  $c_2$  are the mean values inside and outside *C*, respectively.



**Fig. 2** Parameter selection Relative phantom volumes as a function of *k* for phantom volumes 20, 35 and 50 ml and 5 % relative background concentration

For a three-dimensional problem, it can be solved through various methods. One of these methods, developed by Wang et al., is an efficient, iterative convolution-thresholding method (ICTM) for image segmentation. For this study, a 3D extension of the ICTM was applied to the images. This method utilises a number of approximations to establish a framework for minimising energy functionals. A description and mathematical derivation of this approach is beyond the scope of this paper and can be found in the original papers by Wang et al. [28, 33, 34]. It requires only two parameters to be set by the user, a step length ( $\tau$ ) and  $\lambda_1$ , which, in this setting is a function  $\lambda_1(k, \tau)$ . Hence, in practice,  $\lambda_1$  is determined by selecting k and  $\tau$ . The value of  $\tau$  affects the calculation duration, which is not considered in this study, but not whether the calculations converge, as proved by Wang et al. [34]. A value of  $\tau = 0.3$ , within the range used by Wang et el. was used in

**Table 1** Values of k used in the Chan–Vese segmentations, per phantom volume and background AC

Background AC (%)	20 ml	35 ml	50 ml
5	3.10-4	3 · 10 <sup>-4</sup>	3 • 10 <sup>-4</sup>
10	3.10-4	3 · 10 <sup>-2</sup>	3 • 10 <sup>-4</sup>
15	1 • 10 <sup>-2</sup>	6 · 10 <sup>-2</sup>	3 • 10 <sup>-4</sup>
20	2 · 10 <sup>-2</sup>	$5 \cdot 10^{-2}$	3 · 10 <sup>-4</sup>

all calculations. The value of  $\lambda_1(k, \tau)$ , expressed as  $\lambda_1(k, \tau) = k \cdot (\pi \cdot \tau^{-1})^{0.5}$ , does determine whether or not the algorithm converges at a reasonable value. In these calculations, the parameter *k* was selected iteratively for each combination of tunour volume and background AC. Figure 2 shows the relative volume as a function of *k* for a few of the phantoms. The selected values are shown in table 1.

# Results

Figure 3 illustrates segmentation results using the two methods for each of the volumes of the phantoms, using a relative background AC of 10 % and a relative hotspot AC of 200 %. This provides a visual demonstration of how well the segmentation algorithm identifies the thyroid region and separates it from the surrounding background. Both methods describe similar segmentation volumes although slightly different in shape, in this example most prominent for the mid-sized (35 ml) active gland volume. Figure 4 shows the corresponding spatial distribution of misclassified voxels. There does not appear to be any particularly weak areas for either of the methods; the misclassified voxels are distributed across the phantom with no obvious spatial preference.



**Fig. 3** Segmentation example with 200 % hotspot and 10 % background The middle column shows maximum intensity projections of the simulated images for 20 ml (upper), 35 ml(middle) and 50 ml (lower). All images have relative hotspot activity concentration of 200 % and relative background activity concentration of 10 %. The left column shows the corresponding segmentation results from Chan–Vese and the right column shows the results from Otsu



**Fig.** A misclassification example with 200 % hotsport and 10 % background spatial distribution of misclassified voxels displayed as a sum of voxels in the anterior-posterior direction. The middle column shows the true objects in greyscale, where the greyscale level illustrated the number of superimposed object voxels, for 20 ml (upper), 35 ml(middle) and 50 ml (lower). All images have relative hotspot activity concentration of 200 % and relative background activity concentration of 10 %. The left column shows the sum of segmentation errors from Chan–Vese, with missed object voxels in red and misclassified background voxels in blue, and the right column shows the corresponding results from Otsu. These colour brightnesses describe the number of superimposed erroneous voxels. The misclassifications are displayed on a solid gray background of the true object, for orientation purposes

Quantitative results are presented in Figs. 5, 6, 7, 8, 9 and 10, which show the deviation of size estimates for three different phantoms (20 ml, 35 ml, and 50 ml), calculated using the C–V model and Otsu method. The deviation in estimations of size is presented relative to the true size per relative background AC (rows) and relative hotspot AC (columns). The deviation from the true size is presented as a percentage. Both algorithms seem to have a slight tendency to overestimate small volumes and underestimate large ones and the only visually notable difference between the two models is shown for the mid-size (35 ml) phantom whith low (5 % and 10 %) backgrounds: for this phantom and within this background range the C–V model slightly underestimates, whereas the Otsu method slightly overestimates the active volumes while for small gland with high background the situation is the opposite.

The DSC for the two methods, for the same three phantoms, are shown in Figs. 8, 9 and 10. The DSC is presented per relative background AC (rows) and relative hotspot AC (columns).



Percentage deviation (Volume: 20 ml)

Relative hotspot activity (%)

**Fig. 5** Deviation of size estimates for 20 ml phantom. The calculated volumes for the 20 ml phantom, of Chan–Vese and Otsu method, relative to true size per relative background activity cencentration (rows from bottom to top show 5, 10, 15 and 20 %) and relative hotspot activity concentration (columns from left to right show 100, 150, 200 and 250 %). The figures below each panel show the deviation (in percentages) from true size



Percentage deviation (Volume: 35 ml)

**Fig. 6** Deviation of size estimation for 35 ml phantom. The calculated volumes for the 35 ml phantom, of Chan–Vese and Otsu method, relative to true size per relative background activity concentration (rows from bottom to top show 5, 10, 15 and 20 %) and relative hotspot activity concentration (columns from left to right show 100, 150, 200 and 250 %). The figures below each panel show the deviation (in percentages) from true size



## Percentage deviation (Volume: 50 ml)

**Fig. 7** Deviation of size estimates for 50 ml phantom. The calculated volumes for the 50 ml phantom, of Chan–Vese and Otsu method, relative to true size per relative background concentration (rows from bottom to top show 5 %, 10 %, 15 % and 20 %) and relative hotspot activity concentration (columns from left to right show 100 %, 150 %, 200 % and 250 %). The figures below each panel show the deviation (in percentages) from true size



Dice similarity coefficient (Volume: 20 ml)

**Fig. 8** DSC for 20 ml phantom. DSC of Chan–Vese and Otsu method, per relative background activity concentration (rows from bottom to top show 5 %, 10 %, 15 % and 20 %) and relative hotspot activity concentration (columns from left to right show 100 %, 150 %, 200 % and 250 %)



## Dice similarity coefficient (Volume: 35 ml)

**Fig. 9** DSC for 35 ml phantom. DSC of Chan–Vese and Otsu method, per relative background activity concentration (rows from bottom to top show 5 %, 10 %, 15 % and 20 %) and relative hotspot activity concentration (columns from left to right show 100 %, 150 %, 200 % and 250 %)

20 Relative background activity (%) 0.86 0.84 0.84 0.84 0.84 0.83 0.83 0.83 15 Chan Vese 0.9 0.88 0.88 0.87 0.88 0.87 0.87 0.87 True size Otsu 10 0.92 0.91 0.93 0.92 0.9 0.9 0.91 0.9 ß 0.95 0.94 0.93 0.93 0.96 0.95 0.94 0.94 100 150 250 200 Relative hotspot activity (%)

Dice similarity coefficient (Volume: 50 ml)

**Fig. 10** DSC for 50 ml phantom. DSC of Chan–Vese and Otsu method, per relative background activity concentration (rows from bottom to top show 5 %, 10 %, 15 % and 20 %) and relative hotspot activity concentration (columns from left to right show 100 %, 150 %, 200 % and 250 %)



Fig. 11 DSC distribution total distribution of DSC of Chan–Vese and Otsu method



Fig. 12 Background impact DSC as a function of background level for a homogeneous 50 ml phantom

Overall, these figures provide a clear visual representation of the performance of the C-V model and Otsu method in terms of size estimates and segmentation accuracy, under different conditions of relative background AC and relative hotspot AC.

Volume (ml)	MAE (%)	MRV	SD
20	7.6	1.07	0.048
35	4.9	1.01	0.060
50	9.2	0.91	0.070
20-50	7.2	1.00	0.091

 Table 2
 Mean absolute errors, mean relative volumes and standard deviations for the Chan–Vese segmentations, per phantom volume

**Table 3** Mean absolute errors, mean relative volumes and standard deviations for the Otsusegmentations, per phantom volume

Volume (ml)	MAE (%)	MRV	SD
20	11.1	1.07	0.073
35	11.3	1.11	0.021
50	11.1	0.89	0.069
20-50	11.2	1.03	0.059

The total DSC distributions are shown in Fig. 11. A paired *t*-test showed that the difference between the distributions is significant (*p*-value < 0.01), with Chan–Vese being the more accurate algorithm. The variation of DSC between different background levels is shown in Fig. 12.

Tables 2 and 3 present the mean absolute errors, mean relative volumes, and standard deviations of the estimations. In total, the C–V model has slightly lower values for both MAE and MRV than the Otsu method.

Volume	Background AC (%)	Hotspot 0	Hotspot +50	Hotspot +100	Hotspot +150
20	5	80	80	87	92
	10	82	97	97	97
	15	91	91	91	91
	20	107	107	107	107
35	5	103	103	103	103
	10	98	110	110	110
	15	97	97	97	97
	20	89	107	89	107
50	5	80	80	80	80
	10	77	71	77	77
	15	73	73	73	73
	20	80	70	70	70

 Table 4
 Otsu threshold values for each case

## Discussion

C–V model and Otsu method were selected because they are well-tried [19-21] and highly efficient segmentation methods that are not dependent on gradients (sharp edges) [35], which are prone to segmentation image noise and artefacts when input data contain diffuse volume boundaries [36, 37], as is the case in SPECT imaging. (Table 4)

Both methods (Otsu and C–V) manage to segment the images well, and they also tend to deviate in a somewhat similar way from the true volumes. Although not statistically proven, both algorithms seem to have a slight tendency to overestimate small volumes and underestimate large ones.

Although there are some differences in the results, the overall behaviour of the models does not differ substantially. Nevertheless, the, by Wang et al. [28], adapted C–V model performed slightly better and is the more complex of the two, but it requires k to be chosen with care, while Otsu requires no other information than what may be extracted from the image itself. As shown in Fig. 2, the choice of correct parameter value could be a complicated task; this is true in particular for smaller thyroid volumes, as different background activity concentrations are shown to cause different segmentation volume outputs for the same choice of k. Albeit low in complexity, the Otsu method has proved to be a robust segmentation tool [21, 38].Given the similar performance, the simplicity and straightforwardness of the Otsu method make it a powerful competitor.

### Limitations

A major limitation of this study was the relatively low number of simulated cases. In addition to this, the lack of heterogeneity of the thyroid phantoms (in addition to the hotspots) and background AC, as well as the lack of variation in thyroid shape and the shape and composition of the background phantom, as this affects the attenuation properties. Although the physical thyroid phantom contains substantial ammounts of the relatively high-Z components Chlorine  $(_{17}Cl)$  and Sodium  $(_{11}Na)$ , this did not affect the results, as the attenuating properties of the digitised thyroid phantom were set to be water equivalent. Due to the limited resolution of the SPECT images, some voxels will contain image information from playdough (thyroid gland) and water (background). This kind of partial volume effect limits the accuracy of the volume determination. For a thresholding method, such as the Otsu method, the partial volume effect is expected to cause the relative volume overestimation as well as the underestimation of maximum AC being most prominent for the small active volumes [39]. Furthermore, the actual study showed the lowest Otsu segmentation performance, resulting in a volume overestimation as high as 27 %, for the smallest (20 ml) thyroid volume, the highest hotspot AC (250 %) and the lowest background (5 %), i.e. for the highest AC contrast. The explanation for the largest volume overestimation at the highest and not at the lowest contrast is probably the mentioned partial volume effect, prominent for small volumes, combined with the fact that the Otsu method biases towards the image component (either background or foreground) with the largest within-class variance [40, 41]; a lower background in a noisy context, i.e. in SPECT imaging, will have a larger within-class variance than a higher background with the same absolute noise level, as the spectrum will have a lower weight at the lower end for the higher background. This means that the bias will

be the strongest and thus the threshold be lowered the most for the lowest background, resulting in the largest relative volume overestimation for this particular case, in line with the result of the actual study. In conclusion, smaller volumes are harder to segment and the thresholding nature of the Otsu method makes the task of segmenting small volumes more challenging for Otsu compared to the C–V model, the latter not relying on thresholding and thus expected to be less sensitive to partial volume effects.

Since simulated patients do not move at all, these images are free from artefacts, caused by movements or physiological dynamics, which always occur to some extent in vivo. These effects are difficult to quantify but should be of little importance.

## **Future work**

A more thorough investigation of noise level and phantom heterogeneity might be desirable but was considered beyond the scope of this study, as it would be necessary to take methods for homogenisation and noise reduction into consideration to provide optimal conditions for the algorithms. This would be interesting future work, as would a study on the impact of irregularity of phantom shapes, as well as various reconstruction and post-filtration settings. With sufficient access to clinical data, a comparison with AI-based segmentation would also be interesting. With the final objective being the delivery of a correct radiation dose, a dosimetric evaluation of the different methods would be a highly relevant continuation of this work.

# Conclusions

There is a lack and a need for consistency in the determination of thyroid active volume in nuclear medicine images. Even though the, by Wang et al. [28], adapted C–V model drew the longest straw when it comes to accuracy, this study shows a good performance of both methods investigated, under the given circumstances. In terms of autonomy, simplicity and robustness, Otsu is the preferred method. Ultimately it becomes a trade-off between performance and user-friendliness.

#### Abbreviations

AC	Activity concentration
ARF	Angular response function
ATD	Antithyroid drugs
CT	Computed tomography
C–V	Chan–Vese
DSC	Dice similarity coefficient
ICTM 131 <sub>1</sub>	Iterative convolution-thresholding method lodine-131
MAE	Mean absolute error
MLEM	Maximum-likelihood expectation-maximisation
MRV	Mean relative volume
MC	Monte Carlo
MRI	Magnetic resonance imaging
PS	Planar scintigraphy
PET	Positron emission tomography
ROI	Region of interest
SARec	Sahlgrenska academy reconstruction code
SD	Standard deviation
spect 99m <sub>Tc</sub>	Single photon emission tomography Technetium-99 m
T4	Thyroxine

- 3D Three-dimensional
- T3 Triiodothyronin
- 2D Two-dimensional
- US Ultrasound

#### Acknowledgements

We would like to thank Ida Eriksson, medical physicist at Karlstad Central Hospital, for sharing her clinical experience and patiently discussing the project with us, and Henrik Berg, medical physicist at Karlstad Central Hospital, without whose painstaking work in the laboratory during his master's thesis project this paper would never have come to fruition. We are also grateful to Professor Adrian Muntean, at the Department of Mathematics and Computer Science at Karlstad University, for sharing his views on the segmentation methods.

#### Author contributions

JH: Data analysis and manuscript preparation. CA: Data analysis and manuscript preparation. TR: Image generation. JHL: Study conception and design, data preparation, image generation, calculations, data analysis and manuscript preparation. All authors: Commenting on previous versions of the manuscript, reading and approving the final manuscript.

#### Funding

Open access funding provided by Örebro University.

#### Availability of data and materials

The codes used during the current study and the datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

### Declarations

Ethics approval and consent to participate

Not applicable.

#### **Consent for publication**

Not applicable.

### **Competing interests**

The authors have no relevant financial or non-financial interests to disclose.

## Received: 11 April 2023 Accepted: 22 December 2023 Published online: 08 January 2024

#### References

- 1. Maravall F, Gómez-Arnáiz N, Gumá A, Abos R, Soler J, Gomez J. Reference values of thyroid volume in a healthy, noniodine-deficient Spanish population. Horm Metab Res. 2004;36(09):645–9.
- 2. Hall JE, Hall ME. Guyton and hall textbook of medical physiology e-book. London: Elsevier Health Sciences; 2020.
- Bahn RS, Burch HB, Cooper DS, Garber JR, Greenlee MC, Klein I, Laurberg P, McDougall IR, Montori VM, Rivkees SA, Ross DS, Sosa JA, Stan MN. Hyperthyroidism and other causes of thyrotoxicosis: Management guidelines of the American thyroid association and American association of clinical endocrinologists. Endocr Pract. 2011;17(3):456– 520. https://doi.org/10.4158/ep.17.3.456.
- Yip J, Lang BH-H, Lo C-Y. Changing trend in surgical indication and management for graves' disease. Am J Surg. 2012;203(2):162–7. https://doi.org/10.1016/j.amjsurg.2011.01.029.
- Stokkel MPM, Junak DH, Lassmann M, Dietlein M, Luster M. EANM procedure guidelines for therapy of benign thyroid disease. Eur J Nucl Med Mol Imaging. 2010;37(11):2218–28. https://doi.org/10.1007/s00259-010-1536-8.
- Ehlers M. Graves rsquo disease in clinical perspective. Front Biosci. 2019;24(1):35–47. https://doi.org/10.2741/4708.
   Eschner W, Sudbrock F, Weber I, Marx K, Dietlein M, Schicha H, Kobe C. Graves' disease and radioiodine therapy.
- Nuklearmedizin. 2008;47(01):13-7. https://doi.org/10.3413/nukmed-0087.
- Abraham P, Avenell A, Park CM, Watson WA, Bevan JS. A systematic review of drug therapy for graves' hyperthyroidism. Eur J Endocrinol. 2005;153(4):489–98. https://doi.org/10.1530/eje.1.01993.
- Viduetsky A, Herrejon CL. Sonographic evaluation of thyroid size: a review of important measurement parameters. J Diagn Med Sonography. 2019;35(3):206–10.
- Ohiduzzaman M, Khatun R, Reza S, Kadir MA, Akter S, Uddin MF, et al. Thyroid uptake of tc-99m and its agreement with I-131 for evaluation of hyperthyroid function. Univ J Public Health. 2019;7(5):201–6. https://doi.org/10.13189/ ujph.2019.070502.
- 11. Lee WW. Clinical applications of technetium-99m quantitative single-photon emission computed tomography/ computed tomography. Nucl Med Mol Imaging 2019;53(3):172–181. https://doi.org/10.1007/s13139-019-00588-9.
- van Isselt JW, de Klerk JM, van Rijk PP, van Gils AP, Polman LJ, Kamphuis C, Meijer R, Beekman FJ. Comparison of methods for thyroid volume estimation in patients with graves' disease. Eur J Nucl Med Mol Imaging. 2003;30:525–31.
- Hänscheid H, Canzi C, Eschner W, Flux G, Luster M, Strigari L, Lassmann M. EANM dosimetry committee series on standard operational procedures for pre-therapeutic dosimetry II. Dosimetry prior to radioiodine therapy of benign thyroid diseases. Eur J Nucl Med Mol Imaging. 2013;40(7):1126–34. https://doi.org/10.1007/s00259-013-2387-x.

- 14. Varghese J, Rohren E, Guofan X. Radioiodine imaging and treatment in thyroid disorders. Neuroimaging Clin. 2021;31(3):337–44.
- Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, Paschke R, Valcavi R, Vitti P. American association of clinical endocrinologists, American college of endocrinology, and associazione medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules-2016 update appendix. Endocr Pract. 2016;22:1–60.
- Lee H, Kim JH, Kang Y-K, Moon JH, So Y, Lee WW. Quantitative single-photon emission computed tomography/ computed tomography for technetium pertechnetate thyroid uptake measurement. Medicine. 2016;95(27):4170. https://doi.org/10.1097/md.000000000004170.
- 17. Rogowska J. Overview and fundamentals of medical image segmentation. In: Handbook of Medical Imaging, pp. 69–85. Elsevier, San Diego 2000. https://doi.org/10.1016/b978-012077790-7/50009-6.
- 18. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern. 1979;9(1):62-6.
- Husham S, Mustapha A, Mostafa SA, Al-Obaidi MK, Mohammed M, Abdulmaged AI, George ST. Comparative analysis between active contour and otsu thresholding segmentation algorithms in segmenting brain tumor magnetic resonance imaging. J Inform Technol Manag 2020; https://doi.org/10.22059/jitm.2020.78889
- Nazarudin AA, Zulkarnain N, Mokri SS, Zaki WMDW, Hussain A, Ahmad MF, Nordin INAM. Performance analysis of a novel hybrid segmentation method for polycystic ovarian syndrome monitoring. Diagnostics. 2023;13(4):750. https://doi.org/10.3390/diagnostics13040750.
- Gopalakrishnan C, Iyapparaja M. Active contour with modified Otsu method for automatic detection of polycystic ovary syndrome from ultrasound image of ovary. Multimed Tools Appl. 2019;79(23–24):17169–92. https://doi.org/ 10.1007/s11042-019-07762-3.
- 22. Berg H. A phantom based comparison of image segmentation algorithms for adaptive functional volume determination of the thyroid gland using spect. Master's thesis, Stockholm University 2021
- Rydén T, Heydorn Lagerlöf J, Hemmingsson J, Marin I, Svensson J, Båth M, Gjertsson P, Bernhardt P. Fast Gpu-based monte Carlo code for spect/ct reconstructions generates improved 177lu images. EJNMMI Physics. 2018;5(1):1–12.
- Hasinoff SW. Photon, poisson noise. In: Computer vision: a reference guide, pp. 608–610. Springer, Boston, MA(2014. https://doi.org/10.1007/978-0-387-31439-6\_482.
- 25. Cecchin D, Poggiali D, Riccardi L, Turco P, Bui F, Marchi SD. Analytical and experimental FWHM of a gamma camera: theoretical and practical issues. Peer J. 2015;3:722. https://doi.org/10.7717/peerj.722.
- 26. Gonzalez RC, Woods RE. Digital image processing. 3rd ed. New Jersey: Prentice-Hall Inc; 2006.
- 27. MATLAB: Version 9.13.0.2126072 (R2022b Update 3). The MathWorks Inc., Natick, Massachusetts 2022
- Wang D, Li H, Wei X, Wang X-P. An efficient iterative thresholding method for image segmentation. J Comput Phys. 2017;350:657–67.
- 29. Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Process. 2001;10(2):266-77.
- Mumford D, Shah J. Boundary detection by minimizing functionals. In: IEEE Conference on computer vision and pattern recognition, vol. 17, pp. 137–154, 1985. San Francisco
- Mumford D. Shah J. In: Ullman S, Richards W, editors. Boundary detection by minimizing functionals. Berlin: Springer; 1989. p. 19–43.
- Mumford DB, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems. Commun Pure Appl Math. 1989;42(5):577–685.
- Ma J, Wang D, Wang X-P, Yang X. A characteristic function-based algorithm for geodesic active contours. SIAM J Imag Sci. 2021;14(3):1184–205. https://doi.org/10.1137/20m1382817.
- 34. Wang D, Wang X-P. The iterative convolution-thresholding method (ICTM) for image segmentation. Pattern Recogn. 2022;130: 108794.
- Satpute N, Naseem R, Palomar R, Zachariadis O, Gómez-Luna J, Cheikh FA, Olivares J. Fast parallel vessel segmentation. Comput Methods Programs Biomed. 2020;192: 105430. https://doi.org/10.1016/j.cmpb.2020.105430.
- Braiki M, Benzinou A, Nasreddine K, Hymery N. Automatic human dendritic cells segmentation using k-means clustering and Chan-vese active contour model. Comput Methods Programs Biomed. 2020;195: 105520. https://doi. org/10.1016/j.cmpb.2020.105520.
- Seo H, Khuzani MB, Vasudevan V, Huang C, Ren H, Xiao R, Jia X, Xing L. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. Med Phys. 2020. https://doi.org/10.1002/mp.13649.
- Kumar SN, Fred AL, Varghese PS. An overview of segmentation algorithms for the analysis of anomalies on medical images. J Intell Syst. 2020;29(1):612–25.
- Bailey D, Marquis H, Willowson K. Partial volume effect in spect & pet imaging and impact on radionuclide dosimetry estimates. Asia Oceania J Nucl Med Biol 2022; https://doi.org/10.22038/aojnmb.2022.63827.1448
- 40. Xu X, Xu S, Jin L, Song E. Characteristic analysis of Otsu threshold and its applications. Pattern Recogn Lett. 2011;32(7):956–61. https://doi.org/10.1016/j.patrec.2011.01.021.
- Hou Z, Hu Q, Nowinski WL. On minimum variance thresholding. Pattern Recogn Lett. 2006;27(14):1732–43. https:// doi.org/10.1016/j.patrec.2006.04.012.

#### Publisher's Note

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