Computers in Biology and Medicine 120 (2020) 103751

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering

Leonardo Rundo [a,b,*], Lucian Beer [a,b,d], Stephan Ursprung [a,b], Paula Martin-Gonzalez [b,c], Florian Markowetz [b,c], James D. Brenton [b,c], Mireia Crispin-Ortuzar [b,c], Evis Sala [a,b], Ramona Woitek [a,b,d]

[a] Department of Radiology, University of Cambridge, Cambridge CB2 0QQ, UK
[b] Cancer Research UK Cambridge Centre, University of Cambridge, Cambridge CB2 0RE, UK
[c] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK
[d] Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Vienna 1090, Austria

ARTICLE INFO

Keywords:
Tissue-specific segmentation
Computed tomography
Unsupervised fuzzy clustering
Ovarian cancer
Renal cell carcinoma
Radiomics

ABSTRACT

Background: Cancer typically exhibits genotypic and phenotypic heterogeneity, which can have prognostic significance and influence therapy response. Computed Tomography (CT)-based radiomic approaches calculate quantitative features of tumour heterogeneity at a mesoscopic level, regardless of macroscopic areas of hypo-dense (i.e., cystic/necrotic), hyper-dense (i.e., calcified), or intermediately dense (i.e., soft tissue) portions.

Method: With the goal of achieving the automated sub-segmentation of these three tissue types, we present here a two-stage computational framework based on unsupervised Fuzzy C-Means Clustering (FCM) techniques. No existing approach has specifically addressed this task so far. Our tissue-specific image sub-segmentation was tested on ovarian cancer (pelvic/ovarian and omental disease) and renal cell carcinoma CT datasets using both overlap-based and distance-based metrics for evaluation.

Results: On all tested sub-segmentation tasks, our two-stage segmentation approach outperformed conventional segmentation techniques: fixed multi-thresholding, the Otsu method, and automatic cluster number selection heuristics for the K-means clustering algorithm. In addition, experiments showed that the integration of the spatial information into the FCM algorithm generally achieves more accurate segmentation results, whilst the kernelised FCM versions are not beneficial. The best spatial FCM configuration achieved average Dice similarity coefficient values starting from 81.94±4.76 and 83.43±3.81 for hyper-dense and hypo-dense components, respectively, for the investigated sub-segmentation tasks.

Conclusions: The proposed intelligent framework could be readily integrated into clinical research environments and provides robust tools for future radiomic biomarker validation.

1. Introduction

Cancer is typically characterised by genotypic and phenotypic heterogeneity, which has prognostic significance and may influence the response to therapy [1]. Computed Tomography (CT) is the most frequently used cross-sectional imaging method in oncology. It quantifies spatial variation in the morphology of individual tumours by measuring variations in X-ray attenuation, which allows for the assessment of the macro- and mesoscopic structure of tumours [2,3]. Intra- and inter-tumoural heterogeneity can be quantified on the mesoscopic level by using CT-based radiomics, which has been shown to hold both predictive and prognostic information for many cancer types, including high-grade serous ovarian carcinoma (HGSOC) and renal cell carcinoma (RCC). Notably, these two cancer types are characterised by high levels of macroscopic heterogeneity with frequent cystic/necrotic, solid, and calcified tumour regions [3–9]. However, the majority of radiomics studies disregard macroscopic tumour heterogeneity, even though solid tumour regions typically have high cellular density and could contribute more to adverse prognostic or predictive information than necrotic, cystic, or calcified regions [10]. We reasoned

* Corresponding author at: Department of Radiology, University of Cambridge, Cambridge CB2 0QQ, UK.
E-mail addresses: lr495@cam.ac.uk (L. Rundo), lb795@cam.ac.uk (L. Beer), su263@cam.ac.uk (S. Ursprung), paula.martingonzalez@cruk.cam.ac.uk (P. Martin-Gonzalez), florian.markowetz@cruk.cam.ac.uk (F. Markowetz), james.brenton@cruk.cam.ac.uk (J.D. Brenton), mireia.crispinortuzar@cruk.cam.ac.uk (M. Crispin-Ortuzar), es220@cam.ac.uk (E. Sala), rw585@cam.ac.uk (R. Woitek).

https://doi.org/10.1016/j.compbiomed.2020.103751
Received 28 February 2020; Received in revised form 3 April 2020; Accepted 5 April 2020
Available online 10 April 2020
0010-4825/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

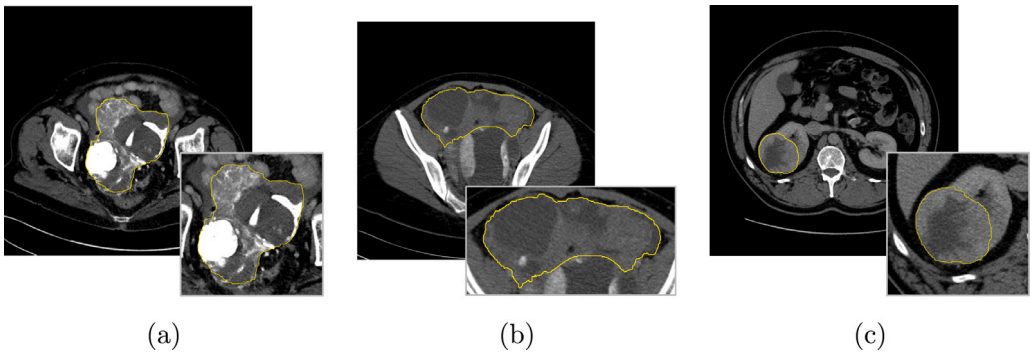

Fig. 1. Examples of input axial CT slices for tissue-specific sub-segmentation: (a, b) HGSOC lesions in the pelvis and omentum, respectively; (c) RCC. The whole tumour burden, defined by the yellow contour and zoomed at the bottom right of each sub-figure, is characterised by mixed tumoural tissues.

that applying different weightings to radiomic features for macroscopically different tumour regions could increase accuracy for predicting response and outcome. However, clinical CT reporting to evaluate the size of tumour masses and response to treatment relies upon mono- or multi-dimensional tumour diameters, typically following RECIST 1.1 criteria [11]. This standard reporting does not quantify the proportion of the tumour that is composed of solid, cystic/necrotic, or calcified tissue [11]. These methods therefore may benefit from automated or manual sub-segmentation of all disease present. Fig. 1 shows three examples of axial CT slices analysed for tissue-specific sub-segmentation of HGSOC and RCC lesions.

Recent advances in machine learning techniques for medical imaging have benefited from the strong learning ability of fully supervised deep learning models [12] and the availability of large training datasets that include accurate and detailed annotations [13,14]. In order to work on datasets with less accurate annotations (for example, bounding boxes or image-level labels [15]), different machine learning models use weak supervision [16] or Generative Adversarial Networks (GANs) [17,18] for data augmentation. In clinical applications, particularly in the case of heterogeneous or multi-institutional datasets, the development of effective supervised deep models typically relies upon solutions tailored for obtaining adequate generalisation abilities, even on limited data samples [19,20]. For this reason, when dealing with an amount of labelled input data that does not allow for a representative training sample along with a sufficient unseen test set, unsupervised machine learning techniques have particularly gained ground in biomedical applications [21,22].

We mainly address the following issue in medical image analysis:

- How accurately does an unsupervised machine learning approach segment hyper-dense and hypo-dense components on the whole tumour burden on CT images?

The rationale underlying this question, towards precision oncology, was:

- May tissue-specific cancer sub-segmentation, as a measure of intra-tumoural heterogeneity, provide insights into a more precise therapy response assessment?

In this work, we propose a computational framework based on unsupervised machine learning techniques to sub-segment tumour lesions into hypo-dense (cystic/necrotic), hyper-dense (calcified), and intermediately dense (soft tissue) tumour components. To the best of our knowledge, this is the first approach that purposely focuses on whole tumour burden sub-segmentation on CT images. Our method optimises the segmentation for each individual image whilst also taking into account prior domain knowledge for the typical densities of candidate sub-regions. Our automated approach allows for deployment in clinical research environments, without the need for any training phase [23]. Furthermore, the results of our tissue-specific sub-segmentation method are interpretable by researchers and clinicians [24,25], by taking into account prior domain knowledge of the typical sub-region Hounsfield Unit (HU) values.

This manuscript is organised as follows. Section 2 concisely introduces the theoretical background of unsupervised fuzzy clustering techniques. The proposed automatic tissue-specific segmentation framework is presented in Section 3. Section 4 presents the characteristics of the analysed HGSOC and RCC datasets, along with the evaluation methodology used. Section 5 shows and discusses the achieved experimental results. Finally, Section 6 provides conclusive remarks and future directions.

2. Unsupervised fuzzy clustering techniques

This section briefly outlines the main concepts underlying the devised unsupervised fuzzy clustering framework designed to unify the classic, spatial, and kernelised versions of the Fuzzy C-Means (FCM) method [26,27]. For a detailed description of the mathematical formulation, please refer to Section S1 in the Supplementary Material.

Fundamentally, the FCM algorithm [26,27] is a partitional clustering technique that minimises the intra-cluster variance, as well as maximises the inter-cluster variance, in terms of a distance metric between the feature vectors [28]. This unsupervised technique optimises the intrinsic partitioning of an unlabelled dataset $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ composed of N feature vectors, which denote data samples $\mathbf{x}_k \in \mathbb{R}^D$ $(k = 1, 2, \dots, N)$ belonging to a D-dimensional Euclidean space, into exactly C clusters (i.e., non-empty partitions of the input dataset). Thus, a partition $\mathcal{P}$ is defined as a fuzzy set family $\mathcal{P} = \{\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_C\}$ [29]. Importantly, let $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_C\}$ be a set of D-dimensional prototype vectors, called centroids that are associated with the C clusters. Therefore, the input dataset $\mathcal{X}$ is partitioned by iteratively searching for the optimal fuzzy partition $\mathcal{P}$ that minimises an objective function $\mathcal{J}_m$ (where m denotes the fuzzification constant) by means of a local optimisation technique. The role of the weighting exponent m in the FCM model was systematically analysed in [30], where the authors suggested that the best choice for m is in the interval $[1.5, 2.5]$, and its mean value $m = 2$ is typically used.

The classic FCM clustering algorithm does not take into account any spatial relationship among pixels since all the samples are used as disperse and independent points, making it sensitive to noise and other imaging artefacts [31]. Accordingly, the integration of spatial information might be beneficial. The spatial FCM (sFCM), elegantly introduced by Chuang et al. [32], enables the retention of the same formulation and objective function as the classic FCM algorithm, just by modifying the update rules with the local spatial content in the image. The incorporation of this spatial component considerably improves the performance: (*i*) in a homogeneous region, the spatial functions emphasise the original membership, so the clustering results are not affected; (*ii*) in noisy regions, spurious blobs or misclassified pixels

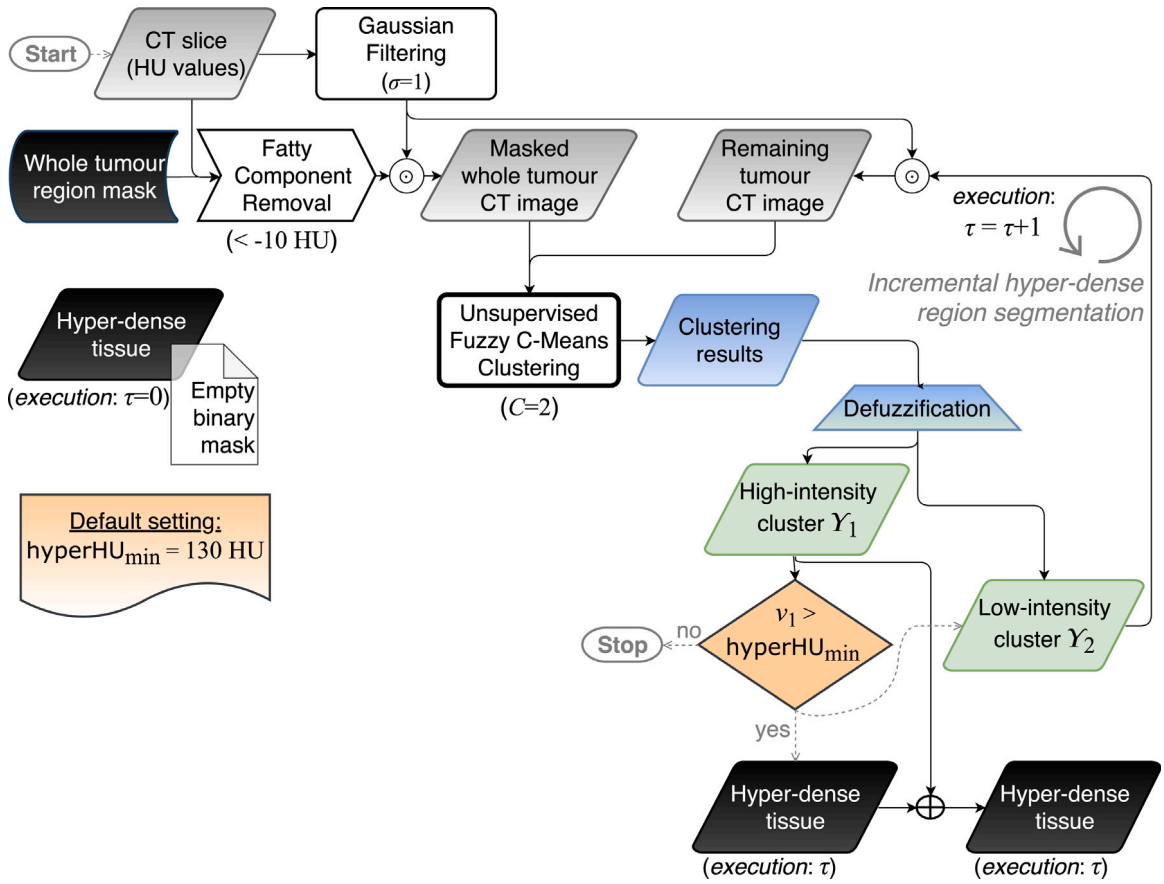

Fig. 2. Flow diagram of the hyper-dense tissue segmentation (i.e., phase 1). The grey and black data blocks denote grey-scale images and binary masks, respectively. The gradient blue–green trapezoidal block represents the defuzzification step, *via* a maximum membership scheme, from the fuzzy clustering results (blue data block) to the crisp clusters (green data blocks). Solid and dashed arrows correspond to processing and control-oriented operations, respectively.

may be corrected. According to [32], in all the tests, a local squared window of $\omega \times \omega$ pixels, with $\omega = 5$, was used. Simply, the parameters p and q weight the original membership (based on pixel values alone) and spatial components, respectively. Hereafter, in compliance with the notation introduced in [32], we denote the sFCM with the control parameters p and q as $\text{sFCM}_{p,q}$. Relying upon the literature [31,32], we tested $p = 1$ and $q \in \{0, 1, 2\}$.

The metric used in the objective function of these FCM versions is still the Euclidean distance. However, the use of the ℓ_2 norm might lead to non-robust results on the segmentation of an image corrupted by noise, outliers, and other imaging artefacts. The kernelised methods let us generalise distance-based algorithms to operate in feature spaces, usually non-linearly related to the input space. Importantly, kernelised methods are suitable for clustering algorithms [33] and also allow for implicit mapping [34]. In our implementation, a Gaussian Radial Basis Function (GRBF) kernel was used:

$$\mathcal{K}(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}, \tag{1}$$

where σ denotes the kernel width. Since σ is a particularly sensitive parameter we relied upon [33], where an adaptive strategy is used to determine the kernel parameters by using the fast bandwidth selection rule in Eq. (2), based on the distance variance of all data points in the collection:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(d_i - \bar{d}\right)^2}, \tag{2}$$

where $d_i = \|\mathbf{x}_i - \bar{\mathbf{x}}\|$ is the distance from the grey-scale of the ith pixel to the grey-scale average of all pixels and $\bar{d}$ is the average of all distances d_i. To perform a comparison independent of centroid initialisation, our kernelised sFCM (ksFCM) version exploited the formulation adopted by the classic FCM algorithm in the original space.

For all the implemented fuzzy clustering methods, the convergence conditions can be defined by comparing the value of the objective function $\mathcal{J}_m$ between two consecutive iterations. The iterative procedure ends when the convergence condition is less than a fixed tolerance value ε_{tol} or the maximum number of allowed iterations $T_{\max}$ is achieved. In all the tests, we used $\varepsilon_{\text{tol}} = 10^{-5}$ and $T_{\max} = 100$.

Regarding the computational complexity (for each iteration), the classic FCM algorithm requires $\mathcal{O}(NCD)$ floating-point operations [35, 36]. With the introduction of the spatial information conveyed by the local $\omega \times \omega$ window, the sFCM version has a time complexity of $\mathcal{O}(NCD + N\omega^2)$. The ksFCM version involves also the kernel distance computation characterised by a quadratic complexity with the number of objects N, resulting in $\mathcal{O}(N^2CD + N\omega^2)$ floating-point operations [36, 37].

In the literature, additional solutions have been proposed to deal with large datasets. Cannon et al. in [38] proposed the approximate FCM to reduce the FCM's time complexity by replacing the exact calculation with approximate ones *via* look-up tables for the Euclidean distances and exponentiation operations. However, these approximations can be relevant mostly for integer-valued data, whilst lead to result quality degradation for real-valued data [35]. In terms of memory reduction, the reformulation of the iterative FCM update steps presented in [35] allows for eliminating the storage of the membership matrix $\mathbf{U} \in \mathbb{R}^{C \times N}$. Nevertheless, our implementation stores this data structure for the membership filtering that considers the spatial neighbourhood for each pixel.

Recent FCM-based techniques mostly aim at improving the search and convergence capabilities of the optimisation process. Careful seeding mechanisms, such as the FCM++ approach [39], adaptively scatter

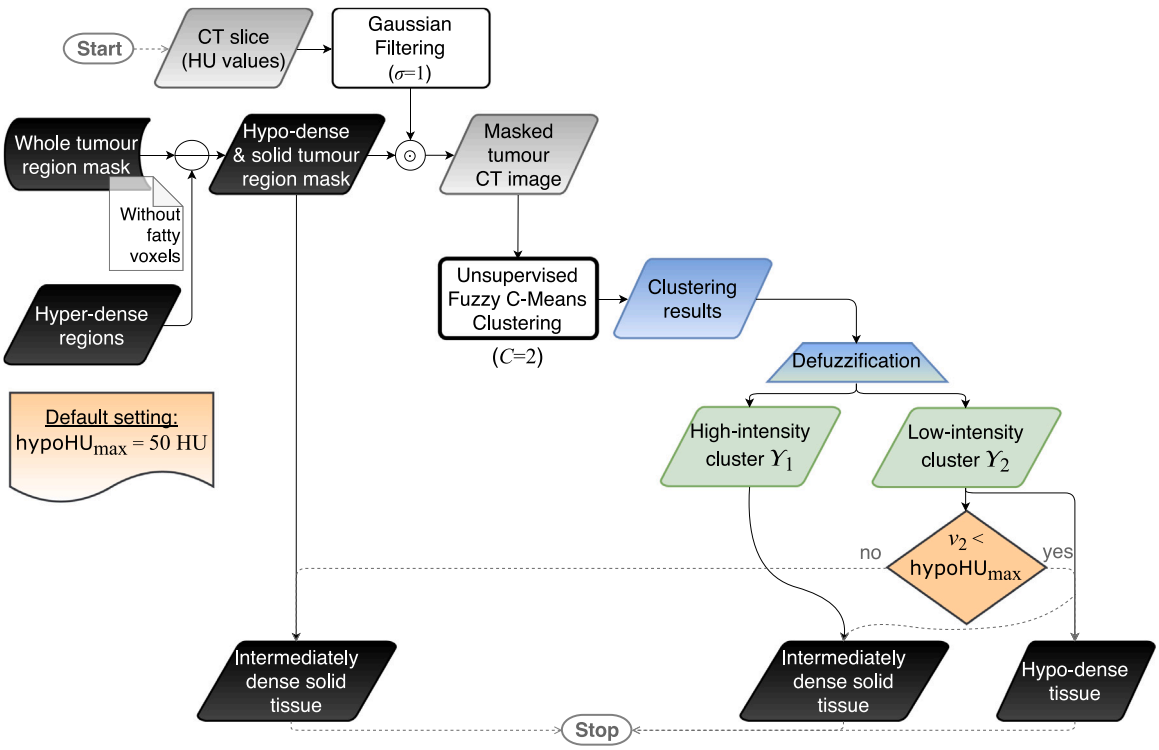

Fig. 3. Flow diagram of the hypo-dense tissue segmentation (i.e., phase 2). The gradient blue–green trapezoidal block represents the defuzzification step, *via* a maximum membership scheme, from the fuzzy clustering results (blue data block) to the crisp clusters (green data blocks). Solid and dashed arrows correspond to processing and control-oriented operations, respectively.

the initial cluster centroids throughout the data space during the initialisation phase. To further boost the FCM performance, extensions and modifications to the objective function can be introduced. For instance, hyper-volume prototypes (with size either fixed or determined automatically from the data undergoing clustering) and heuristic-based adaptive cluster merging or incremental fuzzy partitioning were introduced in [40] and [41], respectively. Alternatively, the search for the optimal solution could be improved by replacing gradient-based search techniques with global optimisation techniques, such as evolutionary strategies [42] or Particle Swarm Optimisation (PSO) [43]. However, these metaheuristics (i.e., population-based stochastic optimisation techniques) are strongly affected by the initialisation of the solutions' population, by influencing both the convergence speed and the quality of the solutions [44,45], as well as careful functioning parameter settings [46]. In this direction, Mekhmoukh and Mokrani in [47] exploited the PSO algorithm for the initial choice of the cluster centroids in brain tissue segmentation on Magnetic Resonance Imaging (MRI) scans. Finally, the fuzzy clustering result was refined *via* level set functions.

In our experiments, the initial fuzzy partitions were randomly generated to carry out a fair comparison independent of centroid initialisation, thus ensuring result repeatability among the unsupervised fuzzy clustering versions investigated in the proposed framework. Moreover, no further computational burden was introduced by careful initialisation schemes.

3. The proposed tissue-specific CT image segmentation method

In our tissue-specific CT image segmentation method, we decided to consider the HU values alone for the segmentation – without including any texture feature (e.g., Haralick features [48,49]) – in order to obtain interpretable results and avoid possible biases in the downstream radiomics analysis (particularly, for feature selection in biomarker development). In this manner, this design choice decouples the morphological tissue-specific sub-segmentation from radiomics-based habitat analyses, as well as maintains the interpretability of the cluster centroids expressed in HU (which are fully understandable for the end-user). Therefore, from now on, the cluster centroids are denoted as scalars $v_i \in \mathcal{V} \subseteq \mathbb{R}^C$.

As a simple pre-processing step, a Gaussian filter (with $\sigma = 1$) was applied by means of a 5×5 convolution kernel. In order to deal with the high bias in the hypo-dense and hyper-dense tissue detection based on unsupervised clustering, a pre-processing step to remove the fatty components was performed; more specifically, the voxels with values lower than -10 HU are removed. This strategy deals with the possible errors in the delineation process (mainly due to the discretisation of the contour drawn by the radiologists that outlines tumours that are surrounded by non-cancerous fat tissue).

The overall sub-segmentation method, relying upon previously delineated whole tumour region masks, leverages a *divide-et-impera* strategy *via* two stages represented in Figs. 2 and 3, respectively:

1. Detection of the hyper-dense regions: multiple executions τ of the unsupervised clustering with $C = 2$, by incrementally including clusters in which the centroid v_1 is higher than the minimum hyper-dense cluster selection value $\mathsf{hyperHU}_{min}$. This iterative procedure takes into account the heterogeneity of the hyper-dense tissues;
2. Distinction between hypo-dense and intermediately dense regions: the clustering algorithm is executed with $C = 2$. Afterwards, the minimum intensity centroid v_2 is compared with the maximum hypo-dense cluster selection value hypoHU_{max}.

This two-stage approach allows us to efficiently avoid the estimation of the number of clusters *via* heuristics, since C is unknown *a priori*. The sequential order of the two phases is motivated by detection purposes of hyper-dense components, which might present small/diffuse calcifications. Besides, the larger HU range of hyper-dense tissues with

respect to hypo-dense portions (even hundreds *versus* few tens in terms of HU value ranges) justifies the choice of multiple executions of the clustering procedure (during phase 1), particularly in the case of highly calcified sub-regions. Afterwards, the delineation of hypo-dense regions can be performed suitably.

To determine the best settings for the cluster selection values, we considered a calibration set consisting of HGSOC lesions containing both hyper-dense and hypo-dense regions. Only two RCC lesions included small calcified areas (see Section 4.1.2) and we used this dataset as an external validation. As a baseline, we used the classic FCM algorithm (i.e., $\text{sFCM}_{1,0}$) without any morphological post-processing to focus on the performance depending only on cluster selection values. The value of $\text{hyperHU}_{\text{min}}$ varied in $\{110, 120, 130, 140, 150\}$ HU considering a calibration set of 70 randomly selected CT images with hyper-dense components. After selecting the best $\text{hyperHU}_{\text{min}}$, the $\text{hypoHU}_{\text{max}}$ values in $\{20, 30, 40, 50, 60\}$ HU were tested on a calibration set composed of 120 randomly selected CT images with hypo-dense components (since hypo-dense tissue is more frequent than hyper-dense regions, as described in Section 4.1). In this study, relying upon the results in supplementary Figs. S1 and S2, the cluster selection values $\text{hyperHU}_{\text{min}}$ and $\text{hypoHU}_{\text{max}}$ were set to 130 HU and 50 HU, respectively, to achieve the best compromise in terms of correct detection performance and reliability, *via* the Dice similarity coefficient (DSC) explained in Section 4.2.2. In more detail, the trend of $\text{hyperHU}_{\text{min}}$ shows a degradation of DSC values for 140 and 150 HU since small calcifications might be overlooked. In the case of $\text{hypoHU}_{\text{max}}$, a value of 20 HU misses the majority of the hypo-dense components, whilst 50 HU shows the lowest standard deviation. Accordingly, we aimed to show the robustness of these settings on unseen data, especially in the case of the RCC dataset. Fig. 4 illustrates the interpretability of our approach *via* a colour-coded HU scale of the different tissues composing the whole tumour burden. Gradient colours were used to show that no fixed threshold can reliably identify the hyper-dense and hypo-dense components. A maximum membership defuzzification scheme was applied after every unsupervised fuzzy clustering execution to yield a crisp classification.

This two-stage approach ensures robustness in highly variable clinical scenarios, such as in the case of metastatic HGSOC that is frequently composed of up to three different tissue types. Using this *divide-et-impera* strategy, no technique for automatic selection of the number of clusters is needed. In fact, the inherent variability within the analysed cohort of patients and tumour types, considering both the different acquisition parameters and tissues occurring in the lesions, might affect the reliability in the estimation of the number of clusters. These strategies include heuristics (e.g., elbow or silhouette methods), information theory methods, or fuzzy clustering validity measures [28,50].

The proposed two-stage approach was applied separately to each lesion type (even when there was more than one distinct lesion in a given CT slice). Furthermore, to increase accuracy, the unsupervised fuzzy clustering was performed independently on each connected-component of the whole tumour. This is important when the regions split or merge across adjacent slices, which is particularly the case in HGSOC.

The proposed approach was developed using the MatLab® R2019b (64-bit version) environment (The MathWorks, Natick, MA, USA). The tests were conducted on a MacOS X (Mojave, version 10.14.6) computational platform equipped with an Intel® Core™ i7@2.7 GHz CPU and 16 GB of RAM.

3.1. Hyper-dense tissue segmentation

Since hyper-dense regions are particularly heterogeneous due to interspersed foci of dense calcifications and non-calcified or less densely calcified tissue (see, for example, Fig. 5), only the sub-regions with the highest HU values would be detected in a single execution of the clustering algorithm. We overcame this problem by performing several executions of the fuzzy clustering with a $C = 2$ procedure using the same cluster selection value $\text{hyperHU}_{\text{min}}$ for each iteration (Fig. 2). More precisely, the clustering algorithm analyses the pixels that were not assigned to the high-intensity cluster $\mathcal{Y}_1$ during the previous iteration until the current υ_1 is lower than $\text{hyperHU}_{\text{min}}$. In this manner, the hyper-dense component is identified by incrementally adding the regions that satisfy the criteria based on the cluster selection value $\text{hyperHU}_{\text{min}}$. Therefore, we can explicitly deal with the heterogeneity of the hyper-dense tissues (i.e., calcifications or vessels). Fig. 5 shows an example of the incremental results achieved by three executions of the clustering procedure. Last, a morphological closing operation (by using a circular structuring element with a two-pixel radius) was applied to make the sub-region boundaries smoother.

3.2. Hypo-dense tissue segmentation

As shown in Fig. 3, the hypo-dense component segmentation relied on the binary mask yielded by the first phase. Indeed, the clustering algorithm was applied on the pixels not assigned to the hyper-dense region (i.e., this binary mask could be also completely 0-valued when no hyper-dense region was previously detected). The hypo-dense and the intermediately dense regions were segmented by using the fuzzy clustering with $C = 2$. After its execution, if the minimum intensity centroid υ_2 was lower than the maximum hypo-dense cluster selection value $\text{hypoHU}_{\text{max}}$, the low-intensity cluster $\mathcal{Y}_2$ was assigned as a hypo-dense region. To achieve a higher sensitivity in the identification of small hypo-dense regions, a fixed thresholding – using the well-established value of 20 HU for cystic/necrotic regions – was employed in the case of no detection *via* the proposed clustering-based pipeline.

Finally, the following morphological operations were performed to refine the sub-segmentation results [51]:

- a small-area removal operation, dealing with connected-components smaller than 0.08 cm^2, to remove small regions not relevant for clinical purposes or radiomic applications;
- morphological closing (circular structuring element with two-pixel radius) to smooth the hypo-dense region boundaries;
- a hole-filling algorithm on the segmented hypo-dense region to remove possible holes due to local inhomogeneities.

4. Materials and evaluation methods

4.1. Patient dataset composition

The proposed framework segments the clinical CT scans of patients affected by (*i*) HGSOC and (*ii*) RCC. All the patients had been referred for clinically indicated CT scans by their clinical team. Both studies were approved by the local ethical review board. Written, informed consent to participate in this research was obtained from patients with ovarian cancer. For patients with RCC, the need for informed consent was waived.

All the analysed CT data are encoded in the 16-bit Digital Imaging and Communications in Medicine (DICOM) format. The dataset comprised axial CT scans acquired at multiple institutions by using scanners from three different vendors: General Electric Healthcare (Waukesha, WI, USA); Siemens Healthineers (Erlangen, Germany); and Toshiba Medical Systems (Tokyo, Japan). The main CT acquisition characteristics for the two datasets are reported in Table 1. Fig. S3 (in Supplementary Material) shows the volume distribution for the whole tumour, hyper-dense and hypo-dense components for the three considered tumour lesion locations. In all the cases, the volume distributions are right-skewed and present outliers, thus showing the intrinsic variability across the samples. Fig. S4 (in Supplementary Material) shows the variability of the Signal-to-Noise-Ratio (SNR), computed as $\text{SNR}_{\text{ROI(WholeTumour)}} = \frac{\mu_{\text{ROI(WholeTumour)}}}{\sigma_{\text{ROI(WholeTumour)}}}$, across the three tumour lesion locations analysed in this study.

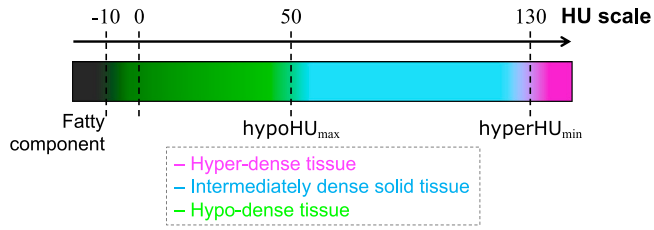

Fig. 4. HU scale showing the different densities of the tissues composing the whole tumour burden on CT imaging. hyperHU$_{min}$ and hypoHU$_{max}$ denote the cluster selection values for the hyper-dense and hypo-dense tissues, respectively. The colour legend is shown at the bottom.

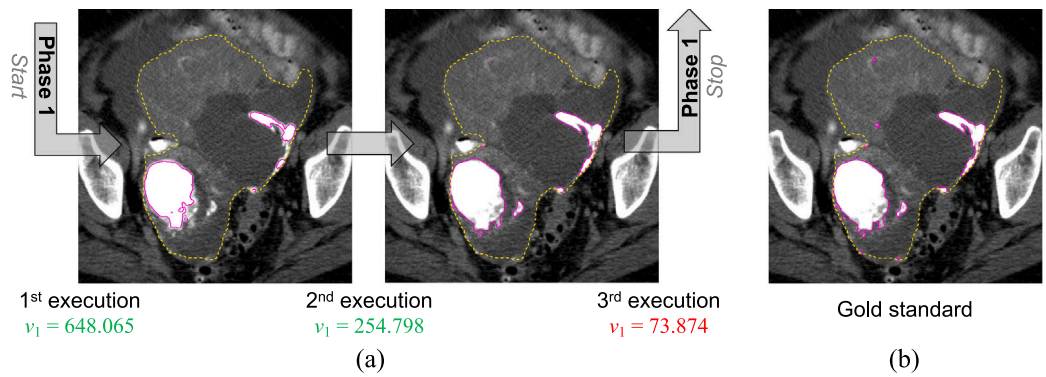

Fig. 5. Incremental hyper-dense region segmentation *via* multiple executions of the unsupervised fuzzy clustering models for hyper-dense region detection. (a) Example of intermediate results obtained by the first phase of the proposed pipeline (employing sFCM$_{1,1}$) on an HGSOC CT image. The high-intensity cluster centroid values v_1, during the executions, are also shown. For better clarity, the green-coloured and red-coloured centroids v_1 denote higher or lower values than the selected cluster selection value hyperHU$_{min}$ = 130, respectively. (b) Corresponding manual gold standard. The whole tumour and the segmented hyper-dense region contours are displayed as dashed yellow and solid magenta lines, respectively.

Table 1
CT acquisition parameters of the HGSOC and RCC datasets.

Dataset	Peak voltage [kV]	Matrix size [pixels]	Slice thickness [mm]	Pixel spacing [mm]
HGSOC	{100, 120, 130}	512 × 512	2.0–5.0	0.627–0.977
RCC	{100, 120, 140}	512 × 512	{3.75, 5.0}	0.586–0.965

4.1.1. High-grade serous ovarian carcinoma

CT scans of the abdomen and pelvis of 29 patients with HGSOC were included in this study. All ovarian cancers contained tumour of intermediate density together with either hypo-dense or hyper-dense portions, or both. We selected the most frequent and clinically relevant anatomic locations of HGSOC metastases, which are the pelvis and ovaries (Pelvic and Ovarian Disease, POD) and in the omentum. Overall, 26 and 10 POD and omental lesions, respectively, were considered. The total number of analysed CT slices was 965, where the average number of slices per lesion was 26.8 ± 19.5 and 25.7 ± 19.4 for POD and omental lesions, respectively. The average lesion volume was highly variable: 769.8 ± 1068.7 cm^3 and 290.1 ± 435.4 cm^3 for POD and omental lesions, respectively. More specifically, considering the tissue-specific Regions of Interest (ROIs), the number of hyper-dense (hypo-dense) regions was 15 (24) and 9 (7) for the POD and omentum, respectively.

4.1.2. Renal cell carcinoma

The RCC dataset was composed of 10 patients with a total number of 152 analysed CT slices (average number of slices per lesion: 15.2 ± 6.2). All the renal lesions considered in this study contained hypo-dense tissue components and only two revealed small calcifications (volume lower than 0.2 cm^3).

The average volume of the lesions was 215.1 ± 182.1 cm^3. Whilst CT scans in patients with HGSOC were acquired during the portal venous phase, renal CT scans were acquired during the nephrographic phase, which involves a longer delay after the injection of intra-venous contrast agent.

4.2. Evaluation methodology

In this section, we describe the gold standard delineation strategy and the used region detection evaluation metrics.

4.2.1. Gold standard delineation procedure

CT images were loaded into Microsoft Radiomics (project InnerEye,[1] Microsoft, Redmond, WA, USA) and the entire POD, as well as any metastases in the omentum, were semi-automatically outlined in consensus by three readers: a medical doctor and Ph.D. student with 1.5 years of training and experience in cancer imaging (S.U.), a radiology registrar with five years of experience (L.B.), and a consultant radiologist with ten years of experience (R.W.) in general radiology and oncological imaging.

Hypo-dense areas that represented cystic or necrotic parts were identified visually and outlined separately. The same was done with the hyper-dense tumour portions that represented calcifications. We optimised window settings for the identification and semi-automatic segmentation of calcified tumour portions similar to the approach proposed in [52]. We measured the attenuation in the solid tumour part by manually placing an ROI$_{solid}$ there. The mean of the HU in the ROI$_{solid}$ was then used to estimate the optimal window level and width, respectively: Window$_{level}$ = HU(ROI$_{solid}$) · 2.68 and Window$_{width}$ = HU(ROI$_{solid}$) · 3.1.

[1] https://www.microsoft.com/en-us/research/project/medical-image-analysis/

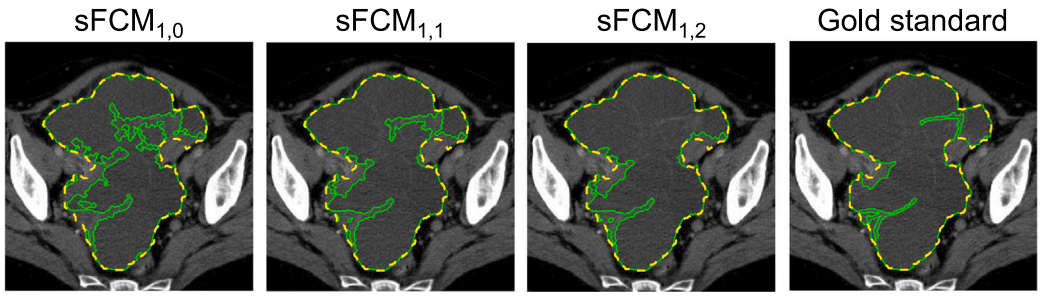

Fig. 6. Influence of the weighting parameters p and q in the $\text{sFCM}_{p,q}$ algorithm compared against the gold standard delineation. The whole tumour, hyper-dense, and hypo-dense region contours are displayed as dashed yellow, solid magenta and solid green lines, respectively.

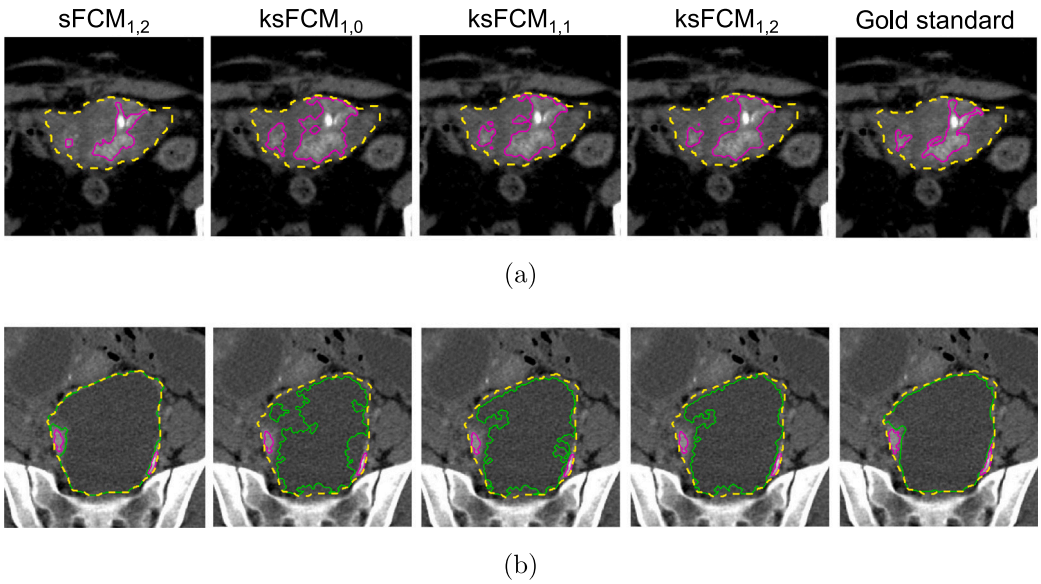

Fig. 7. Example of results achieved by $\text{sFCM}_{1,2}$ compared to $\text{ksFCM}_{1,q}$ by varying the spatial component weight $q \in \{0, 1, 2\}$: (a) hyper-dense tissue segmentation; (b) hypo-dense tissue segmentation. In both cases, the gold standard delineation is shown at the right-most panel. The whole tumour, hyper-dense, and hypo-dense region contours are displayed as dashed yellow, solid magenta and solid green lines, respectively.

4.2.2. Region detection and segmentation evaluation metrics

In order to assess the ability of the proposed method to correctly detect the slices with hypo-dense and hyper-dense components, we calculated the Area Under the Receiver Operating Characteristic Curve (AUROC).

For the quantitative evaluation of the image segmentation results achieved by the investigated computational methods, the automatically segmented CT images (S) were compared against the corresponding gold standard manual segmentation (G) using spatial overlap- and distance-based metrics [53–55]. Since our method analyses 2D CT images (mainly due to the slice thickness that may give rise to disconnected ROIs between adjacent slices), we calculated slice-wise metrics that were then averaged per patient. The segmentation evaluation metrics were computed separately for the hyper-dense and hypo-dense components. To achieve the goal of clinical and radiomic applications, a minimum area of 0.15 cm^2 was set for the sub-region connected-components considered in the segmentation evaluation metrics calculation. In this manner, we decrease the effect on our assessment values caused by potentially created ROIs that consist of too-few pixels to be relevant for clinical or radiomic approaches. The used medical image segmentation evaluation metrics are described in Section S2 of the Supplementary Material.

The two-sided Wilcoxon signed rank test on paired DSC results [56] was performed (for each type of the segmented regions in a slice-wise fashion) with the null hypothesis that the samples come from continuous distributions with equal medians. In all the tests, a significance level of 0.05 was considered.

4.2.3. Competing methods

Since no existing literature work has addressed the tissue-specific sub-segmentation of the whole tumour burden on CT images so far, an experimental comparison of the proposed unsupervised FCM-based techniques was performed against the following segmentation approaches:

- fixed multi-threshold approach, which relies on clinically-established thresholds: pixels with values higher than 220 HU or lower than 20 HU are assigned to the hyper-dense and hypo-dense clusters, respectively. Relying on [57,58], a threshold of 220 HU is generally used for aortic calcifications;
- two-stage Otsu method [59], which executes the same controls, based on the HU values for the inclusion in the hyper-dense and hypo-dense clusters, and post-processing steps;
- automatic selection of the number of regions based on cluster evaluation methods. Considering the crisp K-means clustering algorithm [60], K was estimated for each slice (aiming at a fine-grained control for finding the underlying tissue distribution). The tested heuristics were: the Caliński–Harabasz (CH) criterion [61]; the Davies–Bouldin (DB) criterion [62]; the silhouette criterion [63]; the gap statistics [64]. For all the techniques, the

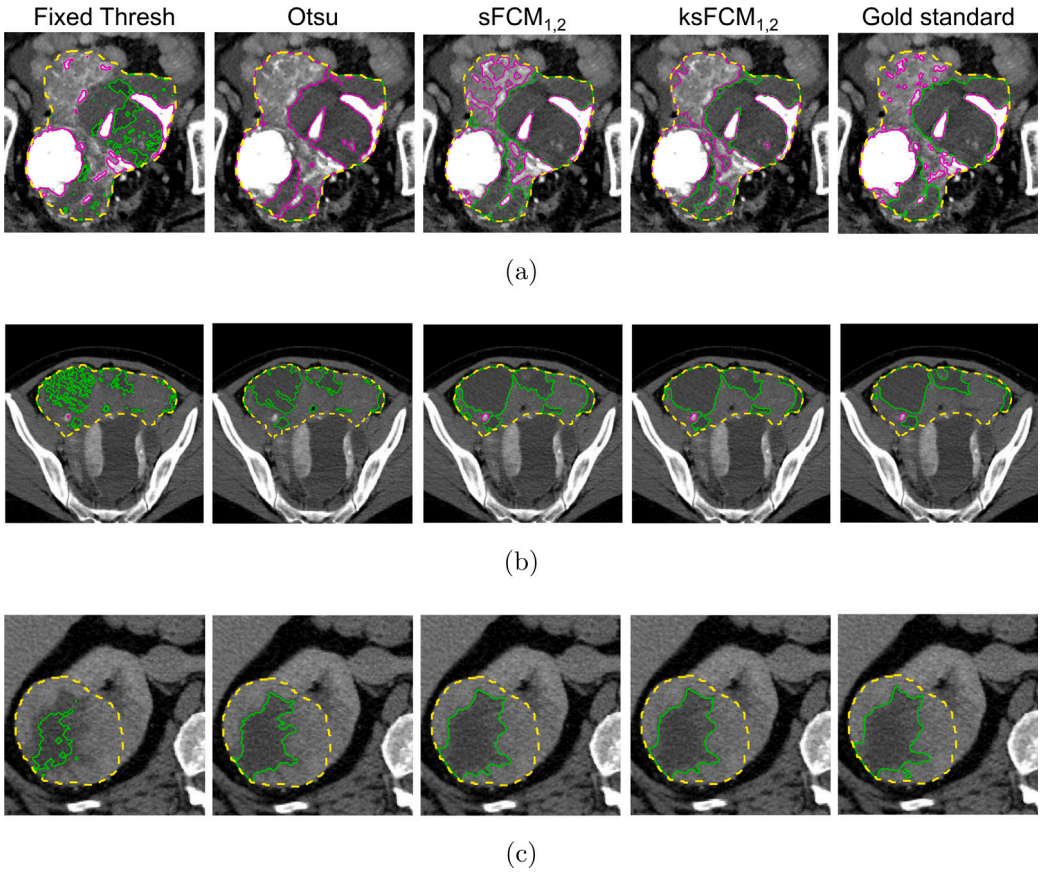

Fig. 8. Segmentation results computed on the input CT images in Fig. 1. The whole tumour, hyper-dense, and hypo-dense region contours are displayed as dashed yellow, solid magenta and solid green lines, respectively.

range of values used was $K \in \{1, 2, 3\}$. The automatic modified FCM cluster segmentation algorithm, proposed by Li and Shen [28], is unsuitable since the used cluster validity function, based on the fuzzy partitions (explicitly considering the cardinality of each cluster), might be highly affected by the ROI sizes and class imbalance.

5. Experimental results

This section presents the experimental results achieved by the proposed computational framework, by showing both graphical examples and quantitative evaluation metrics.

Fig. 6 shows an example of hypo-dense tissue segmentation results by varying the weighting parameters, p and q, in the case of the sFCM algorithm. It is worth noting that the higher the spatial weighting q, the more connected the segmented areas; this applies especially in the case of highly heterogeneous hypo-dense tissue components.

Furthermore, two examples of the implemented ksFCM versions (with various values of the q parameter) compared against the $\text{sFCM}_{1,2}$ for hyper-dense and hypo-dense tissue segmentation are depicted in Figs. 7(a) and 7(b), respectively. In both cases, the introduction of the spatial context also incrementally improves the segmentation results also for the kernelised version. However, the delineations of all the ksFCM are less accurate than those achieved by $\text{sFCM}_{1,2}$.

Fig. 8 shows the results achieved by the implemented methods on the CT images in Fig. 1. For higher visibility, we display only $\text{sFCM}_{1,2}$ and $\text{ksFCM}_{1,2}$ results (achieving the best overall qualitative and quantitative performance among the tested p and q values), along with the fixed thresholding and two-stage Otsu methods. The fixed thresholding, as well as the Otsu method, tends to under-estimate the segmented regions. In particular, in the case of large inhomogeneous hypo-dense components, the segmentation might present many disconnected and spurious areas. In addition, some small calcifications could be missed. Furthermore, the tested two-stage Otsu approach could fail in the case of lesions with highly mixed tissue components (Figs. 8(a) and 8(b)). With regard to unsupervised fuzzy clustering methods, $\text{sFCM}_{1,2}$ generally yields more accurate segmentation results than $\text{ksFCM}_{1,2}$; Fig. 8(a), in particular, shows the high ability to detect diffuse calcified tissue, as well as small details in the hypo-dense component.

To better demonstrate how the tumoural tissue components appear intertwined, Fig. 9 shows three examples of three-dimensional rendering of the ROIs, allowing us to display their actual locations in the whole tumour (represented by means of the enclosing transparent yellow surface).

5.1. Region detection and segmentation results

Table 2 shows the AUROC for evaluating the specificity and sensitivity of the performance of hyper-dense and hypo-dense region detection. The first experimental finding was that the fixed thresholding and the two-stage Otsu method do not perform adequately. Similarly, the automatic strategies for the selection of the number of clusters for the K-means algorithm showed a poor performance. This could be observed particularly in HGSOC due to the higher prevalence of hyper-dense and hypo-dense components, compared to RCC. The proposed two-stage approach based on unsupervised fuzzy clustering achieves excellent detection performance by overcoming the need for the *a priori* number of clusters. On the contrary, the same two-stage approach employing the Otsu method in place of the fuzzy clustering algorithms did not achieve a comparable performance.

Regarding the segmentation evaluation metrics described in Section 4.2.2, for conciseness and clarity, we report only the DSC values in

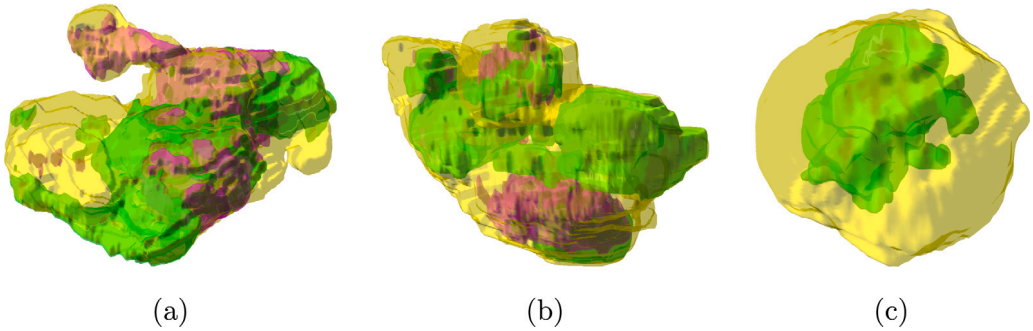

Fig. 9. Three-dimensional reconstruction of the segmented ROIs (green and magenta volumetric models for the hyper- and hypo-dense components, respectively) in their actual location with respect to the enclosing whole tumour (transparent yellow surface): (a) POD, (b) omental lesion, (c) RCC. The transparent surfaces are rendered with alpha blending ($\alpha = 0.40$).

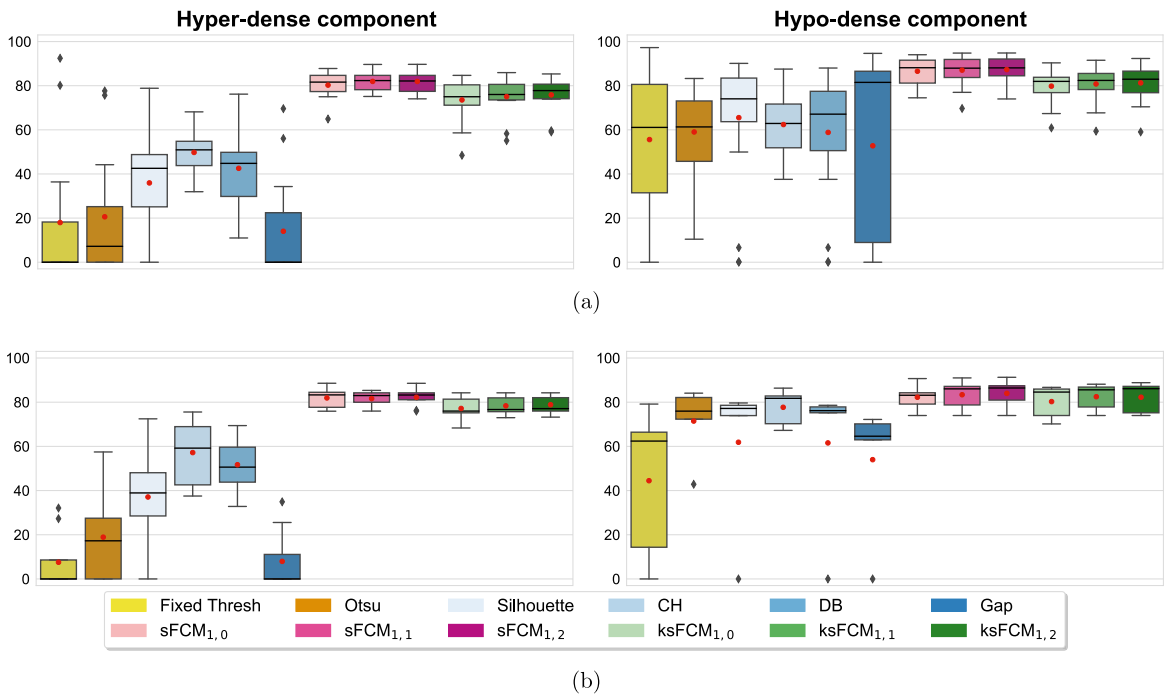

Fig. 10. DSC values of the tissue-specific sub-segmentation results for the (a) POD and (b) omental lesions on the HGSOC CT datasets.

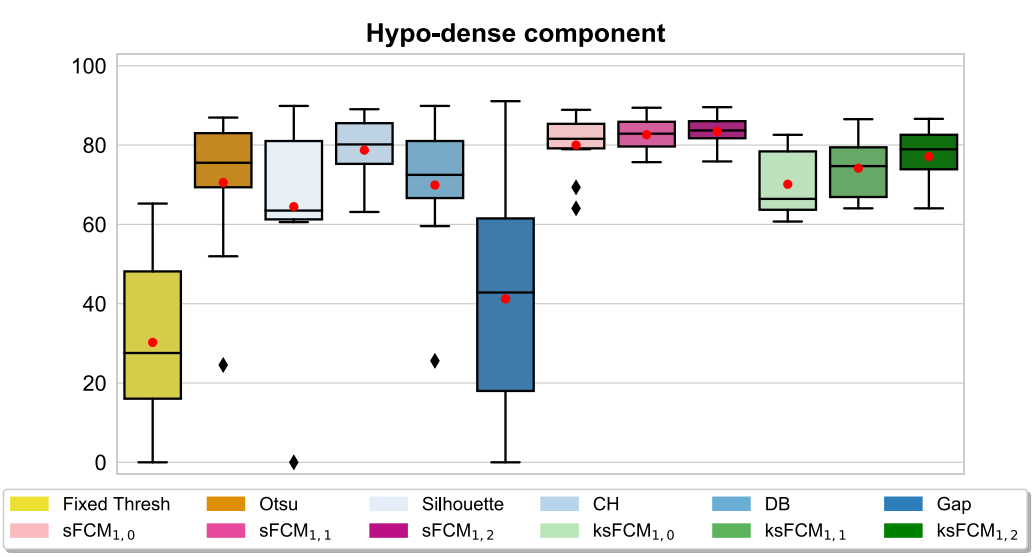

Fig. 11. DSC values of the tissue-specific sub-segmentation results for the kidney lesions on the RCC CT datasets.

Table 2
AUROC achieved by the compared tissue-specific CT image sub-segmentation methods on the HGSOC (POD and omental lesions) and RCC datasets.

Method	HGSOC (POD)		HGSOC (omentum)		RCC	
	Hyper-dense	Hypo-dense	Hyper-dense	Hypo-dense	Hyper-dense	Hypo-dense
Fixed thresholding	0.781	0.534	0.674	0.573	0.667	0.771
Two-stage Otsu	0.682	0.689	0.623	0.716	0.500	0.802
Silhouette + K-means	0.694	0.578	0.572	0.589	0.466	0.739
CH + K-means	0.471	0.501	0.499	0.492	0.333	0.500
DB + K-means	0.481	0.507	0.528	0.535	0.282	0.553
Gap + K-means	0.628	0.754	0.519	0.599	0.490	0.719
$sFCM_{1,0}$	0.901	0.987	0.937	0.987	0.943	0.981
$sFCM_{1,1}$	0.901	0.987	0.937	0.987	0.943	0.981
$sFCM_{1,2}$	0.901	0.987	0.937	0.987	0.943	0.981
$ksFCM_{1,0}$	0.917	0.982	0.937	0.981	0.943	0.981
$ksFCM_{1,1}$	0.901	0.982	0.937	0.981	0.943	0.981
$ksFCM_{1,2}$	0.901	0.987	0.937	0.987	0.943	0.981

what follows. Figs. 10(a), 10(b), and 11 plot the distribution of the DSC values achieved on the POD, omental, and RCC lesions, respectively. All the boxplots display a black solid line and a red circular marker that denote the median and mean values, respectively. The whisker value is set to 1.5 in all cases and the outliers are displayed as black diamonds. The legend box at the bottom denotes the investigated classes of methods with different colour palettes. For completeness, the results of the other metrics are provided in the Supplementary Material (Figs. S5–S14) and are used to support the result analysis.

The fixed thresholding and two-stage Otsu methods obtain low DSC values, because they typically under-estimate the segmented regions (low sensitivity and high specificity values). In accordance with the AUROC values in Table 2, the highly variable DSC results, obtained by the four tested heuristics for the K-means algorithm [60], point out the difficulty of selecting the correct number of clusters; among these strategies, the CH criterion [61] achieved the overall best performance whilst the gap statistics [64] showed highly unreliable results.

In general, the unsupervised fuzzy clustering configuration with $p = 1$ and $q = 2$ outperformed the other configurations for both sFCM and ksFCM. The introduction of the spatial information provided significant benefits over the classic FCM algorithm. However, $sFCM_{1,2}$ overall achieved higher performance than $ksFCM_{1,2}$. In particular, $sFCM_{1,2}$ significantly outperformed $ksFCM_{1,2}$ in the case of POD lesions ($p = 2.282 \times 10^{-4}$ and $p = 1.483 \times 10^{-49}$ for hyper-dense and hypo-dense DSC values, respectively), as well as RCC lesions ($p = 0.0011$) for hypo-dense DSC values, respectively). In more detail, all the ksFCM configurations often fail on noisy images by disconnecting areas with local inhomogeneities, such as large heterogeneous hypo-dense regions. The low presence of hyper-dense components and the large hypo-dense areas in the case of the RCC dataset (Fig. 11), compared to the HGSOC lesions (Figs. 10(a) and 10(b)), can explain the typically better performance of the K-means clustering using the heuristics for the number of cluster selection.

5.2. Computational performance

The computational performance, in terms of processing time and memory consumption, was measured. The execution times were computed by means of the `tic` and `toc` stopwatch timer functions. Moreover, by relying on the size of the variables allocated in the MatLab workspace, we estimated the amount of memory required by the investigated methods. Aiming at a practical use case, we selected a patient with HGSOC in which there was a large pelvic lesion (4690.6 cm^3), with large cystic components, extended across 51 CT slices. (See Table 3.)

The fixed thresholding was the most time-efficient approach, along with the two-stage Otsu method. Among the heuristics for the selection of the number of clusters for the K-means clustering, the silhouette and the gap statistics are particularly demanding with respect to the CH and DB criteria (whose processing times are in line with $sFCM_{1,0}$).

Table 3
Computational performance achieved by the implemented methods in terms of processing times and memory consumption for the HGSOC CT scan (considering a large POD lesion) selected as a practical use case.

Method	Processing time [s]	Memory [GB]
Fixed thresholding	0.1345	1.0185
Two-stage Otsu	1.8046	1.0170
Silhouette + K-means	1.4200×10^3 (= 23.667 mins)	1.0498
CH + K-means	17.9152	1.0498
DB + K-means	18.1542	1.5718
Gap + K-means	2.0796×10^3 (= 34.66 mins)	1.0478
$sFCM_{1,0}$	18.4728	1.0246
$sFCM_{1,1}$	25.3067	1.0246
$sFCM_{1,2}$	24.7146	1.0246
$ksFCM_{1,0}$	31.4829	1.0246
$ksFCM_{1,1}$	52.5809	1.0246
$ksFCM_{1,2}$	51.1077	1.0246

Considering the unsupervised fuzzy clustering implementations, the processing time of sFCM increases with the introduction of the spatial function (see Supplementary Material), even though the computational overhead is mitigated for $q = 1$ and $q = 2$. Comparing $ksFCM_{1,0}$ and $sFCM_{1,0}$, an increase in processing times is appreciable due to the transformation of all the input pixel values into the feature space by means of the GRBF kernel). Interestingly, the trend regarding the spatial version, by varying q in $\{1, 2, 3\}$, is valid also for the kernelised implementations.

With reference to the computational complexity, since our framework uses only the HU values (i.e., $D = 1$) and two clusters (i.e., $C = 2$), the overall time and memory requirements are suitable for nearly real-time performance. Indeed, all the sFCM and ksFCM versions have a linear $\mathcal{O}(N)$ and quadratic $\mathcal{O}(N^2)$ time complexity, respectively. Interestingly, $C = 2$ implies that the membership matrix $\mathbf{U}$ can be stored using $2 \times N$ double-precision floating-point numbers (i.e., 8 bytes in MatLab). Regarding the scalability, since the clustering algorithm is applied for each 2D slice, the implementation scales with the number of slices composing the whole tumour burden. Therefore, a distributed computing paradigm can be leveraged to offload onto multiple CPU cores the independent computations concerning the different slices to segment [65].

6. Discussion and conclusion

In this work, we proposed an intelligent tissue-specific sub-segmentation framework based on unsupervised fuzzy clustering techniques, which allows for clinically interpretable and radiomics-oriented results. Our novel approach, leveraging a two-stage *divide-et-impera* strategy, accurately and efficiently detects and delineates the hyper-dense and hypo-dense components in heterogeneous tumours, thus overcoming the limitations imposed by the automatic selection of the number of clusters required by partitional clustering techniques. We

tested our approach on two datasets comprising highly heterogeneous tumours, namely, HGSOC and RCC. Both detection and segmentation performance with regard to tissue components – in terms of AUROC and overlap-/distance-based evaluation metrics, respectively – showed superiority over the existing methods (namely, fixed thresholding, two-stage Otsu method, automatic clusters number selection heuristics for the K-means clustering algorithm). More specifically, $sFCM_{1,2}$ generally outperformed the other clustering configurations, even when compared to the kernelised versions, in particular. Therefore, the proposed framework could be suitably transferred into biomedical research environments (without requiring any training/set-up phases) for robust radiomic biomarker development [23,66].

From a clinical perspective, the proposed computational framework, yielding interpretable results, might represent a reliable and feasible solution, since it obtains a DSC higher than 70% overall, which is generally regarded as a satisfactory level of agreement between two segmentations (i.e., manual and automated delineations) in clinical applications [67,68]. The accurate segmentation performance achieved by our two-stage framework, in terms of the DSC metric, was confirmed by a good balance of the sensitivity and specificity values. The experimental findings provided by the overlap-based metrics are endorsed by the distance-based metrics that consider the delineated region boundaries. Generally, sFCM achieved more accurate results than ksFCM, consistent with the results presented in [31], where sFCM (with $p = 1$ and $q = 2$) significantly outperformed the K-means, classic FCM, and the kernelised version in brain MRI tissue segmentation.

This single-lesion-focused study on intra-tumoural heterogeneity could be extended to multiple sites to evaluate intra-/inter-tumoural heterogeneity, especially in the case of HGSOC, which typically comprises a heterogeneous mixture of solid and cystic tissue and has frequently metastasised to multiple anatomic locations when diagnosed [69,70]. With regard to RCC, the macroscopic heterogeneity visible on CT is typically caused by necrosis, haemorrhage, and cystic parts [71]. These typical morphological characteristics even allow for a cancer classification based on the appearance of a tumour on CT [72]. CT-based texture feature computation on intermediately dense tumour tissue alone was shown to be effective in the literature: Takahashi et al. [73] drew the largest possible circular ROI avoiding calcifications, whilst Lend et al. [74] manually excluded calcifications and cystic/necrotic parts from the whole tumour. The implementation of our approach for tissue sub-segmentation into a clinical research workflow, which aims at establishing radiomic biomarkers, might allow us to evaluate tissue-type-specific radiomics more extensively against whole-tumour radiomics. Since highly proliferative and aggressive tumour portions are frequently found in solid, non-calcified areas of ovarian cancer [10], radiomics specifically computed for these areas might convey more relevant predictive and prognostic information than global tumour radiomics. Another potential field of application is the clinical radiological setting, where treatment response is commonly assessed based on changes in the overall diameters of tumours [11] whilst this simplification disregards differential changes in solid versus cystic tumour components [75,76]. An automated and reliable approach for the sub-segmentation of tumour sub-regions, as demonstrated here, might allow for more specific response assessment to be first evaluated and subsequently integrated into clinical research environments. Due to the interpretability of the results obtained using our proposed method, clinicians might be more amenable to the implementation of such a tool for clinical purposes compared to less interpretable "black box" approaches [77]. Potential areas of further investigation might regard the integration with circulating biomarkers, where CA125, which is an established clinical biomarker used for disease detection and monitoring in HGSOC [78], as well as circulating tumour cells and cell free tumour DNA in plasma, which are currently evaluated in translational oncological studies [79].

One of the limitations of this study is the continued requirement for relatively labour-intensive and time-consuming manual delineation of tumours and the inherent user-dependence [80]. Convolutional Neural Networks hold the potential to overcome this necessity when exploited to develop a fully automated segmentation approach for combined whole tumour detection and segmentation [81], which could be integrated with our unsupervised tissue-specific sub-segmentation pipeline. However, developing such a comprehensive framework requires large-scale annotated datasets for training/testing and was beyond the scope of this study, but might be a goal for developing the proposed method further. Another limitation is the relatively small number of patients included in this study. However, the large size of some of the selected lesions, which extended over 80 CT slices, also allowed the 2D clustering approach to be validated on a remarkably higher number of images (1117 in total) than the number of patients might suggest. In conclusion, we were able to show the effectiveness of the proposed approach and its advantages compared to the investigated competing methods on both HGSOC and RCC datasets.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.compbiomed.2020.103751.

Acknowledgements

This work was supported by The Mark Foundation for Cancer Research and Cancer Research UK Cambridge Centre [C9685/A25177].

Additional support has been provided by the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Microsoft Radiomics was provided to the Addenbrooke's Hospital (Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK) by the Microsoft InnerEye project.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.103751.

References

[1] R. García-Figueiras, S. Baleato-González, A.R. Padhani, A. Luna-Alcalá, J.A. Vallejo-Casas, E. Sala, et al., How clinical imaging can assess cancer biology, Insights Imaging 10 (1) (2019) 28, http://dx.doi.org/10.1186/s13244-019-0703-0.

[2] J.P. O'Connor, C.J. Rose, J.C. Waterton, R.A. Carano, G.J. Parker, A. Jackson, Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome, Clin. Cancer Res. 21 (2) (2015) 249–257, http://dx.doi.org/10.1158/1078-0432.CCR-14-0990.

[3] H. Lu, M. Arshad, A. Thornton, G. Avesani, P. Cunnea, E. Curry, et al., A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic-and molecular-phenotypes of epithelial ovarian cancer, Nature Commun. 10 (1) (2019) 764, http://dx.doi.org/10.1038/s41467-019-08718-9.

[4] N. McGranahan, C. Swanton, Clonal heterogeneity and tumor evolution: past, present, and the future, Cell 168 (4) (2017) 613–628, http://dx.doi.org/10.1016/j.cell.2017.01.018.

[5] H.A. Vargas, E.P. Huang, Y. Lakhman, J.E. Ippolito, P. Bhosale, V. Mellnick, et al., Radiogenomics of high-grade serous ovarian cancer: multireader multi-institutional study from the Cancer Genome Atlas Ovarian Cancer Imaging Research Group, Radiology 285 (2) (2017) 482–492, http://dx.doi.org/10.1148/radiol.2017161870.

[6] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, V. Buscarino, A. Colarieti, et al., Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months, Eur. Radiol. (2018) 1–11, http://dx.doi.org/10.1007/s00330-018-5389-z.

[7] H. Veeraraghavan, H.A. Vargas, A.J. Sanchez, M. Micco, E. Mema, M. Capanu, et al., Computed tomography measures of inter-site tumor heterogeneity for classifying outcomes in high-grade serous ovarian carcinoma: a retrospective study, bioRxiv, 2019, http://dx.doi.org/10.1101/531046, 531046.

[8] V. Goh, B. Ganeshan, P. Nathan, J.K. Juttla, A. Vinayan, K.A. Miles, Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker, Radiology 261 (1) (2011) 165–171, http://dx.doi.org/10.1148/radiol.11110264.

[9] A. McPherson, A. Roth, E. Laks, T. Masud, A. Bashashati, A.W. Zhang, et al., Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer, Nature Genet. 48 (7) (2016) 758–767, http://dx.doi.org/10.1038/ng.3573.

[10] G. Burkill, S. Allen, R. A'hern, M. Gore, D. King, Significance of tumour calcification in ovarian carcinoma, Br. J. Radiol. 82 (980) (2009) 640–644, http://dx.doi.org/10.1259/bjr/12716831.

[11] E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, et al., New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1), Eur. J. Cancer 45 (2) (2009) 228–247, http://dx.doi.org/10.1016/j.ejca.2008.10.026.

[12] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, http://dx.doi.org/10.1016/j.media.2017.07.005.

[13] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, IEEE J. Biomed. Health Inform. 21 (1) (2017) 4–21, http://dx.doi.org/10.1109/JBHI.2016.2636665.

[14] T. Shaikhina, N.A. Khovanova, Handling limited datasets with neural networks in medical applications: a small-data approach, Artif. Intell. Med. 75 (2017) 51–63, http://dx.doi.org/10.1016/j.artmed.2016.12.003.

[15] M. Rajchl, M.C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, et al., DeepCut: object segmentation from bounding box annotations using convolutional neural networks, IEEE Trans. Med. Imaging 36 (2) (2017) 674–683, http://dx.doi.org/10.1109/TMI.2016.2621185.

[16] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, I.B. Ayed, Constrained-CNN losses for weakly supervised segmentation, Med. Image Anal. 54 (2019) 88–99, http://dx.doi.org/10.1016/j.media.2019.02.009.

[17] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, et al., GAN-Based synthetic brain MR image generation, in: Proc. International Symposium on Biomedical Imaging, ISBI, IEEE, 2018, pp. 734–738, http://dx.doi.org/10.1109/ISBI.2018.8363678.

[18] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, et al., Learning more with less: conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images, in: Proc. 28th ACM International Conference on Information and Knowledge Management, CIKM, ACM, 2019, pp. 119–127, http://dx.doi.org/10.1145/3357384.3357890.

[19] Z. Yan, X. Yang, K.-T. Cheng, Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation, IEEE Trans. Biomed. Eng. 65 (9) (2018) 1912–1923, http://dx.doi.org/10.1109/TBME.2018.2828137.

[20] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, et al., USE-Net: Incorporating squeeze-and-excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets, Neurocomputing 365 (2019) 31–43, http://dx.doi.org/10.1016/j.neucom.2019.07.006.

[21] Z. Zhang, E. Sejdić, Radiological images and machine learning: trends, perspectives, and prospects, Comput. Biol. Med. 108 (2019) 354–370, http://dx.doi.org/10.1016/j.compbiomed.2019.02.017.

[22] I. Aganj, M.G. Harisinghani, R. Weissleder, B. Fischl, Unsupervised medical image segmentation based on the local center of mass, Sci. Rep. 8 (1) (2018) 13012, http://dx.doi.org/10.1038/s41598-018-31333-5.

[23] B. Allen Jr., S.E. Seltzer, C.P. Langlotz, K.P. Dreyer, R.M. Summers, N. Petrick, et al., A road map for translational research on artificial intelligence in medical imaging: From the 2018 national institutes of health/RSNA/ACR/the academy workshop, J. Am. Coll. Radiol. 16 (9, Part A) (2019) 1179–1189, http://dx.doi.org/10.1016/j.jacr.2019.04.014.

[24] A. Vellido, The importance of interpretability and visualization in machine learning for applications in medicine and health care, Neural Comput. Appl. (2019) http://dx.doi.org/10.1007/s00521-019-04051-w.

[25] J.-B. Lamy, B. Sekar, G. Guezennec, J. Bouaud, B. Séroussi, Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach, Artif. Intell. Med. 94 (2019) 42–53, http://dx.doi.org/10.1016/j.artmed.2019.01.001.

[26] J.C. Bezdek, Objective function clustering, in: Pattern Recognition with Fuzzy Objective Function Algorithms, first ed., Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1981, pp. 43–93, http://dx.doi.org/10.1007/978-1-4757-0450-1.

[27] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy C-means clustering algorithm, Comput. Geosci. 10 (2–3) (1984) 191–203, http://dx.doi.org/10.1016/0098-3004(84)90020-7.

[28] Y.-L. Li, Y. Shen, An automatic fuzzy C-means algorithm for image segmentation, Soft Comput. 14 (2) (2010) 123–128, http://dx.doi.org/10.1007/s00500-009-0442-0.

[29] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353, http://dx.doi.org/10.1016/S0019-9958(65)90241-X.

[30] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy C-means model, IEEE Trans. Fuzzy Syst. 3 (3) (1995) 370–379, http://dx.doi.org/10.1109/91.413225.

[31] L. Caponetti, G. Castellano, V. Corsini, MR brain image segmentation: a framework to compare different clustering techniques, Information 8 (4) (2017) 138, http://dx.doi.org/10.3390/info8040138.

[32] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, T.-J. Chen, Fuzzy C-means clustering with spatial information for image segmentation, Comput. Med. Imaging Graph. 30 (1) (2006) 9–15, http://dx.doi.org/10.1016/j.compmedimag.2005.10.001.

[33] M. Gong, Y. Liang, J. Shi, W. Ma, J. Ma, Fuzzy C-means clustering with local information and kernel metric for image segmentation, IEEE Trans. Image Process. 22 (2) (2012) 573–584, http://dx.doi.org/10.1109/TIP.2012.2219547.

[34] S. Chen, D. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, IEEE Trans. Syst. Man Cybern. B 34 (4) (2004) 1907–1916, http://dx.doi.org/10.1109/TSMCB.2004.831165.

[35] J.F. Kolen, T. Hutcheson, Reducing the time complexity of the fuzzy C-means algorithm, IEEE Trans. Fuzzy Syst. 10 (2) (2002) 263–267, http://dx.doi.org/10.1109/91.995126.

[36] T.C. Havens, J.C. Bezdek, C. Leckie, L.O. Hall, M. Palaniswami, Fuzzy C-means algorithms for very large data, IEEE Trans. Fuzzy Syst. 20 (6) (2012) 1130–1146, http://dx.doi.org/10.1109/TFUZZ.2012.2201485.

[37] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, A.K. Nandi, Significantly fast and robust fuzzy C-means clustering algorithm based on morphological reconstruction and membership filtering, IEEE Trans. Fuzzy Syst. 26 (5) (2018) 3027–3041, http://dx.doi.org/10.1109/TFUZZ.2018.2796074.

[38] R.L. Cannon, J.V. Dave, J.C. Bezdek, Efficient implementation of the fuzzy C-means clustering algorithms, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8 (2) (1986) 248–255, http://dx.doi.org/10.1109/TPAMI.1986.4767778.

[39] A. Stetco, X.-J. Zeng, J. Keane, Fuzzy C-means++: Fuzzy C-means with effective seeding initialization, Expert Syst. Appl. 42 (21) (2015) 7541–7548, http://dx.doi.org/10.1016/j.eswa.2015.05.014.

[40] U. Kaymak, M. Setnes, Fuzzy clustering with volume prototypes and adaptive cluster merging, IEEE Trans. Fuzzy Syst. 10 (6) (2002) 705–712, http://dx.doi.org/10.1109/TFUZZ.2002.805901.

[41] R. Scitovski, I. Vidović, D. Bajer, A new fast fuzzy partitioning algorithm, Expert Syst. Appl. 51 (2016) 143–150, http://dx.doi.org/10.1016/j.eswa.2015.12.034.

[42] B. Yuan, G.J. Klir, J.F. Swan -Stone, Evolutionary fuzzy C-means clustering algorithm, in: Proc. International Conference on Fuzzy Systems, FUZZ-IEEE, Vol. 4, IEEE, 1995, pp. 2221–2226, http://dx.doi.org/10.1109/FUZZY.1995.409988.

[43] C. Fuchs, S. Spolaor, M.S. Nobile, U. Kaymak, A swarm intelligence approach to avoid local optima in fuzzy C-means clustering, in: Proc. International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, 2019, pp. 1–6, http://dx.doi.org/10.1109/FUZZ-IEEE.2019.8858940.

[44] P. Cazzaniga, M. Nobile, D. Besozzi, The impact of particles initialization in PSO: parameter estimation as a case in point, in: Proc. Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB, IEEE, 2015, pp. 1–8, http://dx.doi.org/10.1109/CIBCB.2015.7300288.

[45] S. Spolaor, A. Tangherloni, L. Rundo, M.S. Nobile, P. Cazzaniga, Reboot strategies in particle swarm optimization and their impact on parameter estimation of biochemical systems, in: Proc. Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB, IEEE, 2017, pp. 1–8, http://dx.doi.org/10.1109/CIBCB.2017.8058550.

[46] A. Tangherloni, S. Spolaor, P. Cazzaniga, D. Besozzi, L. Rundo, G. Mauri, M.S. Nobile, Biochemical parameter estimation vs. benchmark functions: a comparative study of optimization performance and representation design, Appl. Soft Comput. 81 (2019) 105494, http://dx.doi.org/10.1016/j.asoc.2019.105494.

[47] A. Mekhmoukh, K. Mokrani, Improved fuzzy C-means based particle swarm optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation, Comput. Methods Programs Biomed. 122 (2) (2015) 266–281, http://dx.doi.org/10.1016/j.cmpb.2015.08.001.

[48] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. SMC-3 (6) (1973) 610–621, http://dx.doi.org/10.1109/TSMC.1973.4309314.

[49] L. Rundo, A. Tangherloni, S. Galimberti, P. Cazzaniga, R. Woitek, E. Sala, et al., HaraliCU: GPU-powered haralick feature extraction on medical images exploiting the full dynamics of gray-scale levels, in: V. Malyshkin (Ed.), Parallel Computing Technologies, PaCT, 2019, in: LNCS, vol. 11657, Springer International Publishing, Cham, Switzerland, 2019, pp. 304–318, http://dx.doi.org/10.1007/978-3-030-25636-4_24.

[50] B.N. Li, C.K. Chui, S. Chang, S.H. Ong, Integrating spatial fuzzy Clustering with level set methods for automated medical image segmentation, Comput. Biol. Med. 41 (1) (2011) 1–10, http://dx.doi.org/10.1016/j.compbiomed.2010.10.007.

[51] P. Soille, Morphological Image Analysis: Principles and Applications, second ed., Springer Science & Business Media, Secaucus, NJ, USA, 2004, http://dx.doi.org/10.1007/978-3-662-05088-0.

[52] L. Saba, G. Mallarin, Window settings for the study of calcified carotid plaques with multidetector CT angiography, Am. J. Neuroradiol. 30 (7) (2009) 1445–1450, http://dx.doi.org/10.3174/ajnr.A1509.

[53] A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, BMC Med. Imaging 15 (1) (2015) 29, http://dx.doi.org/10.1186/s12880-015-0068-x.

[54] A. Fenster, B. Chiu, Evaluation of segmentation algorithms for medical imaging, in: Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2005, pp. 7186–7189, http://dx.doi.org/10.1109/IEMBS.2005.1616166.

[55] Y. Zhang, A review of recent evaluation methods for image segmentation, in: Proc. IEEE International Symposium on Signal Processing and its Applications, ISSPA, Vol. 1, IEEE, 2001, pp. 148–151, http://dx.doi.org/10.1109/ISSPA.2001.949797.

[56] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (6) (80–83) 196–202, http://dx.doi.org/10.2307/3001968.

[57] I. Isgum, B. van Ginneken, M. Olree, Automatic detection of calcifications in the aorta from CT scans of the abdomen: 3D computer-aided diagnosis, Academic Radiol. 11 (3) (2004) 247–257, http://dx.doi.org/10.1016/S1076-6332(03)00673-1.

[58] R.V. Buijs, E.L. Leemans, M. Greuter, I.F. Tielliu, C.J. Zeebregts, T.P. Willems, Quantification of abdominal aortic calcification: inherent measurement errors in current computed tomography imaging, PLoS One 13 (2) (2018) e0193419, http://dx.doi.org/10.1371/journal.pone.0193419.

[59] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 11 (285–296) (1975) 23–27, http://dx.doi.org/10.1109/TSMC.1979.4310076.

[60] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inform. Theory 28 (2) (1982) 129–137, http://dx.doi.org/10.1109/TIT.1982.1056489.

[61] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Comm. Stat. 3 (1) (1974) 1–27, http://dx.doi.org/10.1080/03610927408827101.

[62] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (2) (1979) 224–227, http://dx.doi.org/10.1109/TPAMI.1979.4766909.

[63] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, http://dx.doi.org/10.1016/0377-0427(87)90125-7.

[64] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (2) (2001) 411–423, http://dx.doi.org/10.1111/1467-9868.00293.

[65] L. Rundo, A. Tangherloni, P. Cazzaniga, M.S. Nobile, G. Russo, M.C. Gilardi, et al., A novel framework for MR image segmentation and quantification by using MedGA, Comput. Methods Programs Biomed. 176 (2019) 159–172, http://dx.doi.org/10.1016/j.cmpb.2019.04.016.

[66] J.P. O'Connor, E.O. Aboagye, J.E. Adams, H.J. Aerts, S.F. Barrington, A.J. Beer, et al., Imaging biomarker roadmap for cancer studies, Nat. Rev. Clin. Oncol. 14 (3) (2017) 169, http://dx.doi.org/10.1038/nrclinonc.2016.162.

[67] K. Zou, S. Warfield, A. Bharatha, C. Tempany, M. Kaus, S. Haker, et al., Statistical validation of image segmentation quality based on a spatial overlap index, Academic Radiol. 11 (2) (2004) 178–189, http://dx.doi.org/10.1016/S1076-6332(03)00671-8.

[68] A. Zijdenbos, B. Dawant, R. Margolin, A. Palmer, Morphometric analysis of white matter lesions in MR images: method and validation, IEEE Trans. Med. Imaging 13 (4) (1994) 716–724, http://dx.doi.org/10.1109/42.363096.

[69] H.A. Vargas, H. Veeraraghavan, M. Micco, S. Nougaret, Y. Lakhman, A. Meier, et al., A novel representation of inter-site tumour heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome, Eur. Radiol. 27 (9) (2017) 3991–4001, http://dx.doi.org/10.1007/s00330-017-4779-y.

[70] A. Meier, H. Veeraraghavan, S. Nougaret, Y. Lakhman, R. Sosa, R.A. Soslow, et al., Association between CT-texture-derived tumor heterogeneity, outcomes, and BRCA mutation status in patients with high-grade serous ovarian cancer, Abdom. Radiol. 44 (6) (2019) 2040–2047, http://dx.doi.org/10.1007/s00261-018-1840-5.

[71] M.G. Linguraru, J. Yao, R. Gautam, J. Peterson, Z. Li, W.M. Linehan, R.M. Summers, Renal tumor quantification and classification in contrast-enhanced abdominal CT, Pattern Recognit. 42 (6) (2009) 1149–1161, http://dx.doi.org/10.1016/j.patcog.2008.09.018.

[72] T.J. van Oostenbrugge, J.J. Fütterer, P.F. Mulders, Diagnostic imaging for solid renal tumors: a pictorial review, Kidney Cancer 2 (2) (2018) 79–93, http://dx.doi.org/10.3233/KCA-180028.

[73] N. Takahashi, S. Leng, K. Kitajima, D. Gomez-Cardona, P. Thapa, R.E. Carter, et al., Small (< 4 cm) renal masses: differentiation of angiomyolipoma without visible fat from renal cell carcinoma using unenhanced and contrast-enhanced CT, Am. J. Roentgenol. 205 (6) (2015) 1194–1202, http://dx.doi.org/10.2214/AJR.14.14183.

[74] S. Leng, N. Takahashi, D.G. Cardona, K. Kitajima, B. McCollough, Z. Li, et al., Subjective and objective heterogeneity scores for differentiating small renal masses using contrast-enhanced CT, Abdom. Radiol. 42 (5) (2017) 1485–1492, http://dx.doi.org/10.1007/s00261-016-1014-2.

[75] T. Tirkes, M.A. Hollar, M. Tann, M.D. Kohli, F. Akisik, K. Sandrasegaran, Response criteria in oncologic imaging: review of traditional and new criteria, Radiographics 33 (5) (2013) 1323–1341, http://dx.doi.org/10.1148/rg.335125214.

[76] H. Choi, Response evaluation of gastrointestinal stromal tumors, Oncologist 13 (Supplement 2) (2008) 4–7, http://dx.doi.org/10.1634/theoncologist.13-S2-4.

[77] F. Cabitza, R. Rasoini, G.F. Gensini, Unintended consequences of machine learning in medicine, J. Am. Med. Assoc. 318 (6) (2017) 517–518, http://dx.doi.org/10.1001/jama.2017.7797.

[78] N. Scholler, N. Urban, CA125 in ovarian cancer, Biomark. Med. 1 (4) (2007) 513–523, http://dx.doi.org/10.2217/17520363.1.4.513.

[79] F. Mouliere, D. Chandrananda, A.M. Piskorz, E.K. Moore, J. Morris, L.B. Ahlborn, et al., Enhanced detection of circulating tumor DNA by fragment size analysis, Sci. Trans. Med. 10 (466) (2018) eaat4921, http://dx.doi.org/10.1126/scitranslmed.aat4921.

[80] C. Militello, S. Vitabile, L. Rundo, G. Russo, M. Midiri, M.C. Gilardi, A fully automatic 2D segmentation method for uterine fibroid in MRgFUS treatment evaluation, Comput. Biol. Med. 62 (2015) 277–292, http://dx.doi.org/10.1016/j.compbiomed.2015.04.030.

[81] B.D. de Vos, J.M. Wolterink, P.A. de Jong, T. Leiner, M.A. Viergever, I. Išgum, ConvNet-Based localization of anatomical structures in 3-D medical images, IEEE Trans. Med. Imaging 36 (7) (2017) 1470–1481, http://dx.doi.org/10.1109/TMI.2017.2673121.