applied sciences

MDPI

Article

On Unsupervised Methods for Medical Image Segmentation: Investigating Classic Approaches in Breast Cancer DCE-MRI

Carmelo Militello [1,*], Andrea Ranieri [2], Leonardo Rundo [3,4,5], Ildebrando D'Angelo [6,7], Franco Marinozzi [2], Tommaso Vincenzo Bartolotta [6,8], Fabiano Bini [2] and Giorgio Russo [1]

1 Institute of Molecular Bioimaging and Physiology, Italian National Research Council (IBFM-CNR), 90015 Cefalu, Italy; giorgio.russo@ibfm.cnr.it
2 Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, 00184 Roma, Italy; ranieri.1837003@studenti.uniroma1.it (A.R.); franco.marinozzi@uniroma1.it (F.M.); fabiano.bini@uniroma1.it (F.B.)
3 Department of Radiology, University of Cambridge, Cambridge CB2 0QQ, UK; lr495@cam.ac.uk
4 Cancer Research UK Cambridge Centre, Cambridge CB2 0RE, UK
5 Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, 84084 Fisciano, Italy
6 Department of Radiology, Fondazione Istituto "G. Giglio", 90015 Cefalu, Italy; ildebrando.dangelo@hsrgiglio.it (I.D.); tommasovincenzo.bartolotta@unipa.it (T.V.B.)
7 Breast Unit, Fondazione Istituto "G. Giglio", 90015 Cefalu, Italy
8 Section of Radiology—Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND), University Hospital "Paolo Giaccone", 90127 Palermo, Italy
* Correspondence: carmelo.militello@ibfm.cnr.it

Abstract: Unsupervised segmentation techniques, which do not require labeled data for training and can be more easily integrated into the clinical routine, represent a valid solution especially from a clinical feasibility perspective. Indeed, large-scale annotated datasets are not always available, undermining their immediate implementation and use in the clinic. Breast cancer is the most common cause of cancer death in women worldwide. In this study, breast lesion delineation in Dynamic Contrast Enhanced MRI (DCE-MRI) series was addressed by means of four popular unsupervised segmentation approaches: Split-and-Merge combined with Region Growing (SMRG), k-means, Fuzzy C-Means (FCM), and spatial FCM (sFCM). They represent well-established pattern recognition techniques that are still widely used in clinical research. Starting from the basic versions of these segmentation approaches, during our analysis, we identified the shortcomings of each of them, proposing improved versions, as well as developing ad hoc pre- and post-processing steps. The obtained experimental results, in terms of area-based—namely, Dice Index (DI), Jaccard Index (JI), Sensitivity, Specificity, False Positive Ratio (FPR), False Negative Ratio (FNR)—and distance-based metrics—Mean Absolute Distance (MAD), Maximum Distance (MaxD), Hausdorff Distance (HD)—encourage the use of unsupervised machine learning techniques in medical image segmentation. In particular, fuzzy clustering approaches (namely, FCM and sFCM) achieved the best performance. In fact, for area-based metrics, they obtained DI = 78.23% ± 6.50 (sFCM), JI = 65.90% ± 8.14 (sFCM), sensitivity = 77.84% ± 8.72 (FCM), specificity = 87.10% ± 8.24 (sFCM), FPR = 0.14 ± 0.12 (sFCM), and FNR = 0.22 ± 0.09 (sFCM). Concerning distance-based metrics, they obtained MAD = 1.37 ± 0.90 (sFCM), MaxD = 4.04 ± 2.87 (sFCM), and HD = 2.21 ± 0.43 (FCM). These experimental findings suggest that further research would be useful for advanced fuzzy logic techniques specifically tailored to medical image segmentation.

Keywords: medical image segmentation; breast cancer; pattern recognition; machine learning; clinical feasibility; magnetic resonance imaging; computer-assisted segmentation

check for updates

Citation: Militello, C.; Ranieri, A.; Rundo, L.; D'Angelo, I.; Marinozzi, F.; Bartolotta, T.V.; Bini, F.; Russo, G. On Unsupervised Methods for Medical Image Segmentation: Investigating Classic Approaches in Breast Cancer DCE-MRI. *Appl. Sci.* **2022**, *12*, 162. https://doi.org/10.3390/app12010162

Academic Editor: Syoji Kobashi

Received: 10 November 2021
Accepted: 19 December 2021
Published: 24 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The use of advanced imaging technologies has significantly improved the quality of medical care delivered to patients, allowing medical imaging to be an essential part of today's healthcare system [1]. In fact, medical imaging comprises techniques for acquiring

images that convey detailed information about the anatomy and physiology of the imaged organs [2]. Moreover, many imaging-enabled tools were developed, supporting clinicians in several tasks of the care process: assisted segmentation [3,4], diagnosis support [5], treatment response assessment [6], radiomic analyses [7,8]. For these reasons, computer-assisted image analysis is considered an essential instrument in the clinical workflow [9,10].

Despite the technological progress characterizing the modern era, some medical tasks, such as image annotation, are still performed manually, often via time-consuming and operator-dependent procedures. As an example, the Gross Tumor Volume (GTV) segmentation for radiotherapy treatments is usually delineated by means of a fully manual procedure [11,12]. Considering that dozens of slices have to be contoured, this manual process is extremely time-consuming. Moreover, operator dependence is critical in terms of result reproducibility. Indeed, these manual procedures are strongly dependent on clinician's knowledge and experience: this means that a remarkable intra- and inter-operator variability can seriously affect the segmentation and quantification results.

In this scenario, computer-assisted approaches (automatic or semi-automatic) allow us to mitigate some of the typical drawbacks of manual procedures. Semi-automatic segmentation techniques involve a minimal level of user interactions and exploit automated algorithms to produce accurate and repeatable results. User interaction, for example, may involve the selection of an approximate initial ROI, which is subsequently used to segment the image [13]. As a consequence, semi-automatic approaches provide more reproducible measurements—compared to fully-manual ones—with a significant reduction of the segmentation time. In particular, fully-automatic segmentation approaches do not require any user interaction. Most existing fully-automatic approaches exploit machine learning or deep learning techniques—such as Support Vector Machines (SVMs) or deep Convolutional Neural Networks (CNNs)—to successfully handle the variability characterizing biomedical data [14,15]. In this scenario, supervised learning techniques are, in general, more complex since they require high computation times and a large amount of labeled data for training. Furthermore, it is important to point out that—from a clinical feasibility perspective—machine learning approaches, which do not require training, are advantageous: the amount of labeled data needed to adequately train and evaluate the approaches based on deep learning is not always available [16,17].

The aim of this work is to show the potential of unsupervised pattern recognition techniques, which do not require training and can be more easily integrated into care routine, especially from a clinical feasibility perspective. For this reason, supervised approaches—such as CNNs—were not treated here. As a relevant case study, we consider the segmentation of contrast-enhancing masses on DCE-MRI. It is worth noting that, in [18], classical unsupervised techniques, both automatic and semi-automatic, allowed us to obtain results comparable or superior to the deep learning approaches.

From experimental evidence, it is possible to observe that fuzzy clustering techniques significantly outperformed direct region detection approaches (i.e., split-and-merge and region growing) and crisp k-means. FCM and sFCM obtained comparable results, although the integration of spatial information into the sFCM allowed for the best performance. Therefore, the explicit management of segmentation uncertainty via multiple degrees of class memberships, along with spatial information, represented the best computational framework for the problem at hand. The main contributions of this study are:

- unsupervised segmentation methods, based on classic pattern recognition techniques, can still provide an effective solution despite the prevalence of supervised approaches, such as deep CNNs;
- the use of traditional unsupervised approaches—requiring no training—is an advantage in terms of of immediate clinical feasibility;
- fuzzy clustering techniques significantly outperformed direct region detection approaches and crisp k-means;
- the integration of spatial information into the FCM algorithm achieved the best performance;

- with the goal of providing a guide for beginners, as well as possibly enabling new future extensions from other researchers, this study provides all the technical information needed to understand both the functioning of each of the studied algorithms and the implemented workflow.

The remainder of this work is organized as follows. Section 2 introduces the theoretical background about unsupervised segmentation approaches, focusing on the algorithms used in this study. A detailed description of the performed analysis and of the implemented processing steps is proposed in Section 3, where the exploited DCE-MRI dataset is also described. Section 4 formulates the metrics used to evaluate the performance of the analyzed approaches. Section 5 illustrates the experimental results, including a discussion about the comparison of the proposed techniques. Finally, discussion and conclusions are provided in Section 6.

2. Theoretical Background

A number of algorithms and techniques for image segmentation have been developed and implemented over the years, and a large amount of literature papers about this non-trivial task were proposed. The aim of this section is to present a brief overview about theoretical notions of literature approaches from which this work drawn inspiration. For this reason, we decided for a comprehensive description, to provide the reader with all the technical information needed to understand the functioning of each of the investigated algorithms, as well as the implemented workflow.

The segmentation involves the image partitioning into homogeneous and meaningful sub-regions. By a formal point of view, the segmentation of an image $\mathcal{I}$ involves the identification of a finite set of regions $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_N$ as in Equation (1):

$$\bigcup_{i=1}^{N} \mathcal{R}_i = \mathcal{I}' \tag{1}$$

with the following constraints:

$$\mathcal{R}_i \cap \mathcal{R}_j = \emptyset, \quad \text{for} \quad i \neq j, \tag{2}$$

$$\mathsf{P}(\mathcal{R}_i) = \mathsf{TRUE}, \quad \text{for} \quad i = 1, 2, \ldots, N, \tag{3}$$

$$\mathsf{P}(\mathcal{R}_i \cup \mathcal{R}_j) = \mathsf{FALSE}, \quad \text{for} \quad i \neq j. \tag{4}$$

With more details, P is an appropriate logical predicate leading the segmentation process. Equation (1) states that the union of all the sub-regions resulting from segmentation process. Equation (2) points out that the intersection of two different sub-regions is the empty set: this means that the segmented sub-regions do not overlap each other. According to Equation (3), the result of the logical predicate P on all the pixels belonging to the same sub-region is always TRUE; in other words, all the pixels belonging to the same region share the same characteristics according to the predicate P. As a consequence, Equation (4) states that the result of P on the union of two distinct sub-regions is FALSE.

It is important to clarify that each type of medical image has a specific set of features reflecting its own properties: in fact, each image is the result of a complex interaction between the human body and the scanner (i.e., X-rays for CT, magnetic fields for MRI, radioactive decay for nuclear medicine exams) [19]. As a consequence, not all the segmentation techniques obtain the same results on all image types, but some algorithms yield better results when applied on a specific kind of image. Furthermore, it is necessary to point out that, typically, an approach that works very well with one type of image does not mean that it continues to perform well even on different images. As a matter of fact, bioimages exhibit a very high variability, as they depend on various factors, both intrinsic (e.g., patients) and extrinsic (e.g., imaging modalities, acquisition parameters). Accordingly,

ad hoc modifications might be required at the level of both the segmentation approach and the pre-/post-processing phases.

In what follows, we outline the techniques investigated and compared for contrast-enhancing mass segmentation on DCE-MRI.

2.1. Split-and-Merge Combined with Region Growing

The simplest segmentation approaches use a global threshold applied on pixel intensity to partition the original image: pixels with an intensity greater than threshold T are assigned to one region, while those below the threshold T are assigned to another one [20–23]. In this way, a binary image is created providing the segmentation of the original image with respect to the chosen threshold value. Split-and-Merge and Region Growing (SMRG) is basically a threshold-based approach that combines Split-and-Merge (SM)—composed of a first split (top-down) phase of the image followed by a merge (bottom-up) phase—with RG for the refinement of the identified regions [23].

The SM algorithm represents a valid alternative to thresholding, because it can find homogeneous regions in terms of uniformity criteria [24–27]. Unlike SM, the RG algorithm, starting from one or more seed-points, identifies an ROI through a growing procedure guided by appropriate similarity properties that describe ROI intensity features. Generally SR and RG are used individually for image segmentation, but the use of both together allows us to exploit the overall potential [28,29].

With more details, the idea behind the SM algorithm involves successive splits of the whole image into disjoint regions lead by a logical predicate P. The algorithm starts with an arbitrary partition $\mathcal{R}$ of the original image (i.e., the whole image) and yields an output composed of uniform sub-regions $\mathcal{R}_i$, for $i = 1, 2, \ldots, n$, according to the logical criterion expressed by P. At the generic step t, if $\mathsf{P}(\mathcal{R}_i) = \mathsf{FALSE}$, each region $\mathcal{R}_i$ is split into four sub-regions (also called 'quad-regions'): this process iteratively continues until a quad-region such that $\mathsf{P}(\mathcal{R}_i) = \mathsf{TRUE}$ or with an area smaller than a certain threshold is found. The logical predicate chosen for this work allows us to find the quad-regions with a mean intensity that is greater than the threshold value yielded by the Otsu's method [30]. If only splitting is used, the final partition contains adjacent regions with identical properties: this drawback can be overcome by merging only adjacent regions where the combined pixels satisfy the predicate P.

After the initial rough ROI identification obtained by means of SM, the RG algorithm expands this initial region to properly identify the lesion. According to the classic RG algorithm, each region begins its own growth from a single pixel (seed-point). Instead, when the SM is exploited, the seed-point can be obtained by using the ROI yielded by the SM algorithm (more correctly, a seed-region) [23]: the seed-region is iteratively grown by evaluating, for each pixel on the boundary, its 8-neighborhood as candidate for the growth. A stopping rule is necessary for interrupting the growing procedure if no more pixels match the membership criterion, which often refers to the proximity of the pixel intensities.

2.2. K-Means

Classification refers to data labeling into disjoint sets according to a common set of features. Among these, clustering algorithms can be used to determine the natural structures in the data. Clustering algorithms can use more sophisticated properties of the image: in digital imaging, this means that spatial and/or spectral features concerning pixels can be exploited [31].

K-means is an unsupervised clustering technique that aims at partitioning an input set of N observations into k clusters [32]. Let $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ be a set of vector observations such that $x_i \in \mathbb{R}^n$ for $i = 1, 2, \ldots, N$. In image segmentation, each component of a vector x represents a numerical pixel attribute: if segmentation is based on gray-scale intensity alone, the n-dimensional observation $x_i \in \mathbb{R}^n$ degenerates into the scalar value $x_i \in \mathbb{R}$ representing the intensity of the i-th pixel. The final purpose of k-means is to partition the N observations into $k < N$ disjoint cluster sets $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k\}$ such that the sum of the distances from each point in a set to the mean of that set is minimum. From

a mathematical point of view, k-means turns classification into an optimization problem with the following cost function in Equation (5):

$$\mathcal{J}(\mathcal{X}, \mathcal{V}) = \underset{c}{\operatorname{argmin}}\Big(\sum_{i=1}^{k}\sum_{x\in\mathcal{C}_i}|x - v_i|^2\Big), \tag{5}$$

where v_i is the centroid of the samples in the set $\mathcal{C}_i$ for $i = 1, 2, \ldots, k$. The function $\mathcal{J}(\mathcal{X}, \mathcal{V})$ has no analytical solution: as a result, k-means proceeds by iteratively finding the minimum of its cost function. In particular, at each iteration t the centroid values of the prototype set $\hat{\mathcal{V}}^{(t)}$) are updated according to Equation (6).

$$\hat{v}_i^{(t)} = \frac{1}{|\mathcal{C}_i^{(t)}|}\sum_{x\in\mathcal{C}_i} x', \tag{6}$$

where $|\mathcal{C}_i^{(t)}|$ is the number of objects belonging to the i-th at the step t.

2.3. Fuzzy C-Means

Biomedical images are characterized by an intrinsic uncertainty, thus causing not well-defined regions (e.g., blurry boundaries or poor anatomical details). This aspect makes thresholding approaches and crisp approaches (such as k-means) not always suitable. Therefore, the natural fuzziness characterizing the Fuzzy C-Means (FCM) clustering approach allows to reach better segmentation results than the hard partitioning offered by k-means [33–36].

The FCM algorithm is an unsupervised clustering technique that searches for the optimal partition of an input data set. The idea leading the FCM classification process is that of minimizing the intra-cluster variance as well as maximizing the inter-cluster variance, in terms of a distance metrics between the feature vectors. Formally, the FCM technique searches for the optimal partition of an input data set $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ of N objects into C clusters. With respect to the k-means algorithm in which each point is assigned to one cluster only, FCM allows each object to belong to multiple clusters with different degrees of membership. This *'soft'* classification allows us to define a fuzzy partition, $\mathcal{P}$ defined as a fuzzy set family $\mathcal{P} = \{Y_1, Y_2, \ldots, Y_C\}$ such that each point can have a partial membership to multiple clusters. In mathematical terms, the matrix $U = [u_{ik}] \in R^{C\times N}$ denotes a fuzzy C-partition of the data set X by means of C membership functions $u_i : X \to [0, 1]$, whose values $u_{ik} := u_i(x_k) \in [0, 1]$ represent membership grades of each element x_k to the i-th fuzzy cluster Y_i, and have to hold the constraints in Equation (7):

$$\begin{cases} 0 \le u_{ik} \le 1 \\ \sum\limits_{i=1}^{C} u_{ik} = 1, \forall k \in 1, 2, \ldots, N \\ 0 < \sum\limits_{i=1}^{N} u_{ik} < N, \forall i \in 1, 2, \ldots, C \end{cases}. \tag{7}$$

Although hard clustering works well on compact and well-separated groups of data, in many real-world situations clusters overlap each other: as a result, assigning them with gradual memberships by exploiting a soft computing approach may be more appropriate. Computationally, the FCM algorithm assigns to the sample x_k the membership function values using the relative distance (i.e., intensity value similarity) of x_k from the C prototype points $\mathcal{V} = \{v_1, v_2, \ldots, v_C\}$ identifying the centroids of the C clusters. Such as k-means algorithm, FCM may be rewritten as an optimization problem with respect to the objective function in Equation (8).

$$\mathcal{J}_m(U, \mathcal{V}; \mathcal{X}) = \sum_{i=1}^{C}\sum_{k=1}^{N}(u_{ik})^m|x_k - v_i|^2, \tag{8}$$

where:

- m is the fuzzification constant (i.e., a weighting exponent such that $1 \leq m < \infty$) controlling the fuzziness of the classification process. If $m = 1$, the FCM algorithm degenerates to a k-means clustering: in general, the higher the m value the greater will be the fuzziness degree (the most common value is $m = 2$);
- U is the fuzzy C-partition of the set $\mathcal{X}$;
- $|x_k - v_i|^2$ is the Euclidean distance between the elements x_k and the centroid $v_i \in \mathcal{V}$.

Considering that the optimization problem described by FCM does not have a closed-form solution, the minimum of the cost function $\mathcal{J}_m(U, \mathcal{V}; \mathcal{X})$ has to be found iteratively. In particular, at each iteration t, the centroid values of the prototype set $\hat{\mathcal{V}}^{(t)}$ and the elements of the matrix $\hat{U}^{(t)}$ are updated according to Equations (9) and (10), respectively.

$$\hat{v}_i^{(t)} = \frac{\sum_{j=1}^{N} (\hat{u}_{ij}^{(t)})^m x_j}{\sum_{j=1}^{N} (\hat{u}_{ij}^{(t)})^m}, \tag{9}$$

$$\hat{u}_{ik}^{(t)} = \left(\sum_{j=1}^{C} \left(\frac{|x_k - \hat{v}_i^{(t)}|}{|x_k - \hat{v}_j^{(t)}|} \right)^{\frac{2}{m-1}}, \right)^{-1}, \tag{10}$$

with $m > 1$ and $x_k \neq \hat{v}_j^{(t)}, \forall j, k$.

At each iteration, each object x_k is compared with the elements of the centroid vector and is assigned to the nearest cluster. The process stops when the convergence condition (i.e., the matrix norm distance between $\hat{U}^{(t+1)}$ and $\hat{U}^{(t)}$ is less than a fixed value (i.e., minimum improvement in the objective function $\mathcal{J}$ between two consecutive iterations ϵ) or the maximum number of iterations $T_{\max}$ is reached. After the convergence, a defuzzification is applied to assign each pixel to the cluster with the highest membership degree, thus achieving a binary classification.

2.4. Spatial Fuzzy C-Means

The traditional FCM clustering does not take into account spatial relationship among neighboring pixels, making it sensitive to noise and other imaging artifacts [37,38]. Breast lesions generally tend to grow in an isotropic way, preserving a pseudo-spherical appearance [39]. Relying on those features, it is expected that neighbouring pixels in a digital image are highly correlated and that the probability that they belong to the same cluster is great. Therefore, the use of the sFCM, taking advantage of spatial relationship of neighbouring pixels, can help image segmentation [40,41]. The spatial function used by sFCM is defined in Equation (11):

$$h_{ij} = \sum_{k \in \mathcal{N}(x_j)} u_{ik}, \tag{11}$$

where $\mathcal{N}(x_j)$ represents a square neighborhood (in the spatial domain) around the pixel x_j. The term h_{ij} represents the probability that the pixel x_j belongs to i-th cluster: as a result, the spatial function of a pixel for a cluster is large if the majority of its neighbourhood belongs to the same cluster. The contribution of the spatial function modifies the classic FCM membership function according to Equation (12):

$$u'_{ij} = \frac{u_{ij}^p h_{ij}^q}{\sum_{k=1}^{C} u_{kj}^p h_{kj}^q}, \tag{12}$$

where p and q control the relative importance of both functions. In a homogeneous region, the spatial functions simply emphasizes the original membership and the clustering remains

unchanged. On the other hand, for a noisy pixel this formula reduces the weighting of a noisy cluster by the labels of its neighboring pixels: as a result, misclassified pixels from noisy regions can easily be corrected. The sFCM clustering process involves two steps at each iteration: the former, which is the same as that in standard FCM, allows for calculating the membership function in the features domain, while the latter maps the membership information of each pixel into the spatial domain and allows to calculate the spatial functions for each pixel of the image. At this point, the FCM iteration proceeds with the new membership that is incorporated with the spatial function and stops when the maximum difference between cluster centers at two successive iterations is lower than a certain threshold. After the convergence, a defuzzification scheme (i.e., maximum membership) is applied.

3. Materials and Methods

As introduced in Section 1, the purpose of this work is to present an in-depth analysis of classical unsupervised segmentation algorithms, developing them appropriately to adapt them to the clinical case addressed (i.e., breast lesions detection), thus improving the overall performance.

This section, along with showing the characteristics of the DCE-MRI dataset analyzed, describes the processing pipeline of the proposed analysis. It is necessary to point out that particular attention was paid to the optimization of each step for the specific case study, basically on three different levels:

- *pre-processing*: obtaining images with similar characteristics for the downstream processing steps;
- *segmentation*: optimization of the parameters of the investigated segmentation algorithms;
- *post-processing*: definition of appropriate region properties (based on connected-components) to refine the segmentation results.

3.1. MRI Dataset Description

The analysis of this study was performed on a clinical DCE-MRI dataset composed of 50 patients with breast cancer: a total of 599 slices were processed. The main details on MRI acquisition parameters are reported in Table 1 , while in Figure 1 the phases related to some benign and malignant lesions are shown. The dataset includes patients with different stages of breast cancer, allowing us to cover a wide clinical scenario: as matter of fact, this dataset contains various levels of segmentation difficulty with some scans showing low contrast and large inhomogeneities. Lesions with a non-homogeneous enhancement region, masses with an irregular shape or necrotic core are also included.

Table 1. Some characteristics of the DCE-MRI dataset used for this work.

Characteristic	Value
Scanner manufacturer	General Electric (GE)
Scanner model	Signa HDxt
MRI sequence	DCE-MRI
Time Repetition [ms]	37.720
Time Echo [ms]	17.640
Slice thickness [mm]	2
Interslice spacing [mm]	1
Slice pixel spacing [mm]	0.6875
Matrix size [pixels]	512×512

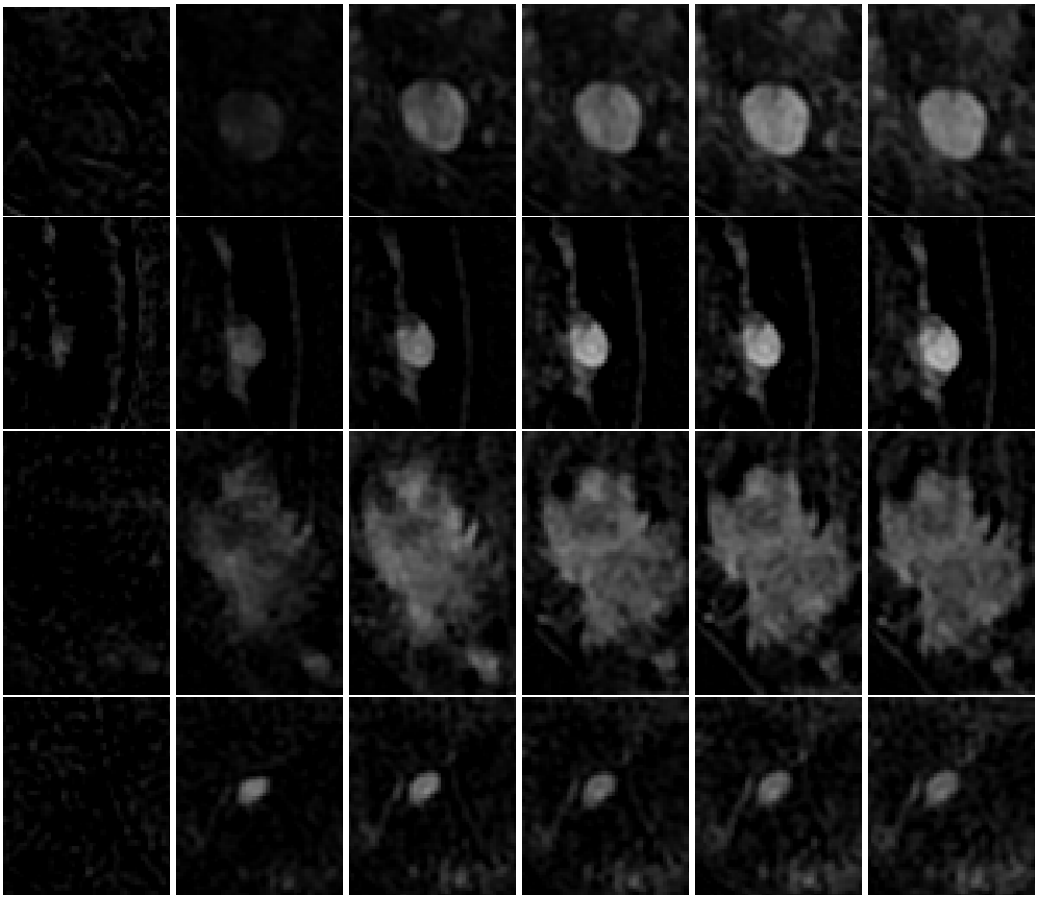

Figure 1. Examples of DCE-MRI phases related to four breast cancer patients (one per row): malignant lesions are shown on rows 1 and 3, while benign lesions are shown on rows 2 and 4. Among phases, for segmentation purposes the clinician selected the strongest one. In particular, for these lesions, the 4th, 3rd, 4th, and 3rd phase was chosen, respectively.

3.2. The Proposed Analysis

Clinical knowledge points out that breast lesions in DCE-MRI appear hyperintense compared to the adipose and muscle surrounding tissues. Furthermore, anatomical atlases and diagnosis reports also refer that breast lesions have the tendency to grow in an isotropic manner, preserving a pseudo-spherical shape [39]. Relying on these morphological features, in this study we decided to start the segmentation from the central slice, which should be the one with the largest area: adjacent slices were processed successively. The idea behind this choice is to segment initially the slice in which the lesion appears more evident, finding the centroid of the connected-component identifying the lesion. On slices different from the central one, the lesion ROI was found by identifying the connected-component with the centroid that has the minimum distance from the centroid found for the central slice. The whole segmentation process can be summarized as follows:

- *dataset loading*: all DCE-MRI images (belonging to the selected patient) and its ground-truth are loaded;
- *ROI selection*: a bounding box containing the lesion is manually drawn in the central image: in this way, the algorithm analyzes only the pixels within the box, reducing computational times and avoiding classification mismatches;
- *pre-processing*: necessary to reduce noise and provide images with similar characteristics for the next processing steps;
- *lesion segmentation*: once the images are pre-processed, the MR image stack of each patient is segmented by all the analyzed methods (i.e., SMRG, k-means, FCM, and sFCM);
- *post-processing*: performed to refine the segmentation and to properly identify the ROI from each binary mask obtained in the previous step;

- *performance evaluation*: the final masks are stored and performance metrics (i.e., area-based and distance-based metrics) are calculated comparing the masks against the ground-truth.

In addition to the well-known segmentation techniques, also the exploited pre- and post-processing operations are widely used by the international community in the field of medical imaging. In fact, the literature offers many works that use the same pre-processing [42,43] and post-processing [44,45] steps in the image analysis phases (before segmentation), as well as in the refinement phases (after segmentation), with excellent results.

Figure 2 shows the flow diagram of the implemented and proposed analysis. Each processing block is described in the following subsections. All the methods were developed using the MatLab environment (Natick, MA, USA). The code is available via GitHub: https://github.com/carmilitello/UnsupervisedSegmentation.git (accessed on 9 November 2021).

3.2.1. Dataset Loading and ROI Selection

This step selects the MRI series to be analyzed. The data loading step, even though does not represent a real processing step, is essential for each algorithm—both supervised and unsupervised—to obtain the data to process. Moreover, the concept of 'supervision' refers generally to the training phase of an algorithm by requiring labeled data. For these reasons, the 'unsupervised' nature of the algorithm used here is preserved. In order to reduce computational times and improve segmentation performance, a minimal user interaction is needed. In fact, by manually tracing a rectangular bounding-box on the image, the operator provides to approximately select the initial ROI containing the breast lesion [13]. Once traced, this ROI is used to crop all the slices of the MR image stack

The manual selection of the initial ROI containing the lesion is useful to reduce the processing time (since only a portion of the whole image is processed), as well as to exclude regions containing pixels with characteristics similar to the lesion that could complicate the algorithm performance, thus invalidating the final result. Moreover, the initial, interactive input provides the clinician with the ability of confidently controlling the entire segmentation process, which is generally preferred compared to a fully automatic process [18,46]. Moreover, considering that the method works on an ROI—selected by the clinician—it is assumed that the tumor is always present within the ROI.

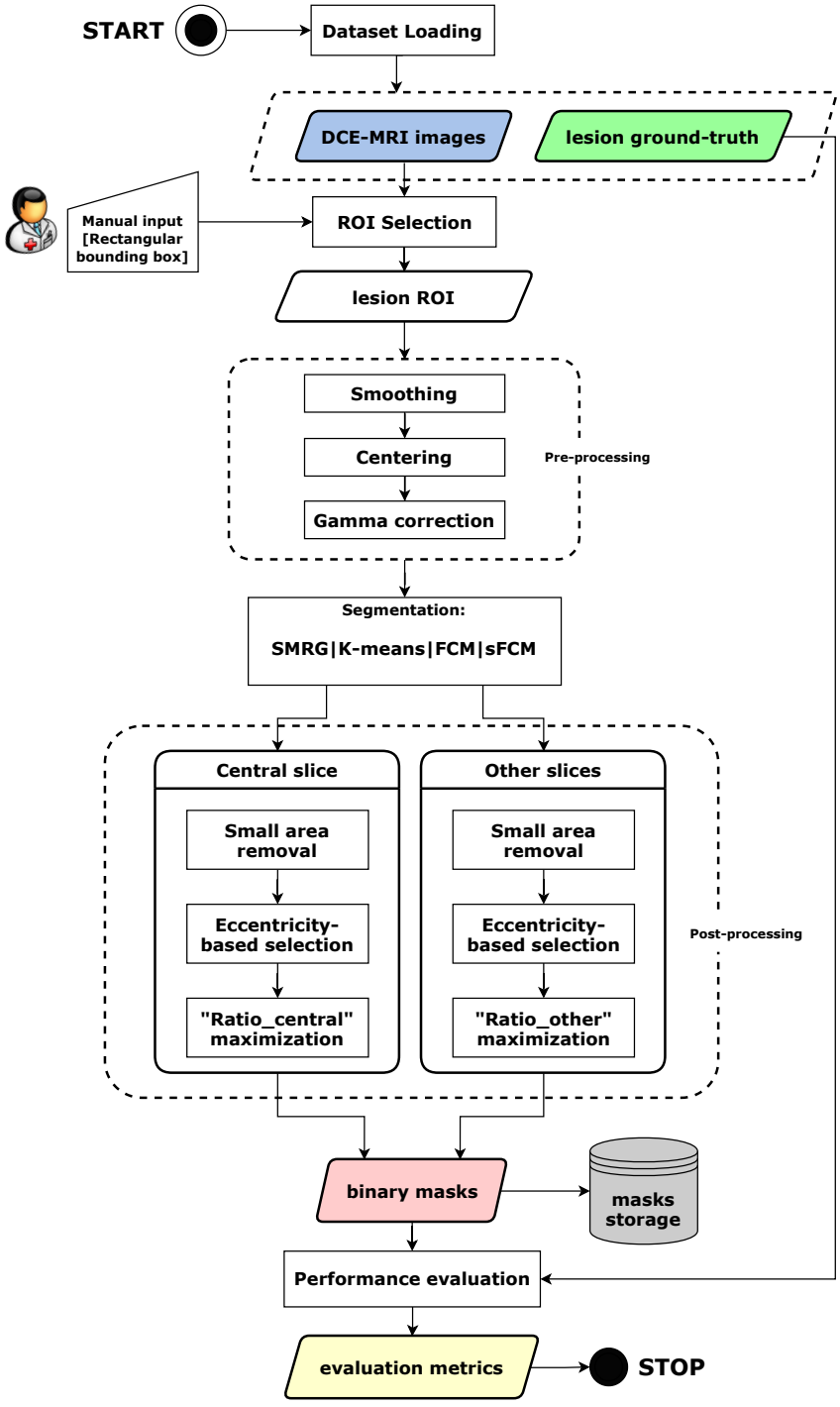

Figure 2. Flow diagram of the proposed analysis for breast lesion segmentation.

3.2.2. Pre-Processing

In order to improve segmentation results, some pre-processing operations are applied after the ROI selection (Figure 3). In particular, the aim of this preliminary phase is to perform denoising and data pre-processing, allowing us to achieve reliable results during the segmentation phase.

The first pre-processing operation deals with noise reduction [47]. The use of median or average (with Gaussian or flat kernels) filters is a common choice in MR image processing. For instance, median filtering was applied to facilitate the ROI identification by reducing the outlier introduced by anatomical peculiarities [48] or also deal with small patient shifts [49]. We used Gaussian kernels having the form in Equation (13), which are the only circular

symmetric kernels that are also separable and that allow you to reduce noise by altering the image less than the average filter.

$$G(r) = ke^{-\frac{r^2}{2\sigma^2}}, \tag{13}$$

where r, σ and k represent the radius, the standard deviation and the normalization factor of the Gaussian function G, respectively. Kernel normalization, obtained by multiplying its coefficients by k—obtained as the inverse of the sum of all kernel coefficients—has two purposes: *(i)* the average value of an area of constant intensity would equal that intensity in the filtered image (as it should) and *(ii)* it prevents the introduction of biases during filtering (i.e., the sum of the pixels in the original and filtered images will be the same). For the purpose of this work, a Gaussian kernel with $r = 5$ and $\sigma = 0.6$ was chosen: this standard deviation value performed denoising without an excessive blurring of the original image.

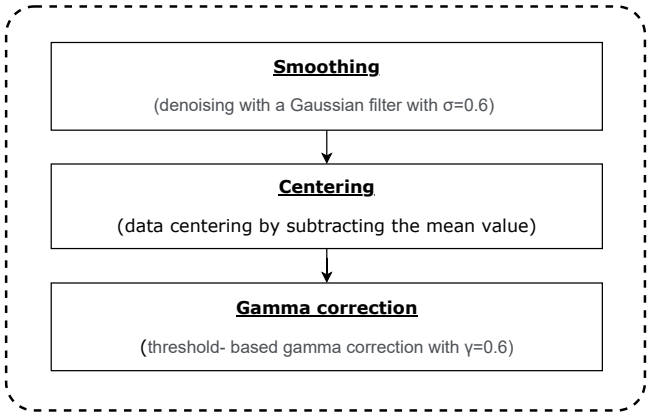

Figure 3. Flow diagram of the pre-processing steps approach.

The next pre-processing step subtracts the mean value from the original image. This is a very helpful step in signal processing because it emphasizes signal variations by shifting its mean to zero. For this reason, during the pre-processing phase, the algorithm removed this bias from the original image amplifying gray level variations. Once the mean value subtraction, the pixels resulting in a negative value have to be clipped to zero in order to avoid visualization problems in the following steps.

After noise reduction and mean subtraction, the last step performs a modified gamma-transformation to stretch the original image histogram. The general form of a gamma transformation is defined in Equation (14):

$$s = cr^{\gamma}, \tag{14}$$

where c and γ are positive constants. Power-law curves with fractional values of γ map a narrow range of dark input values into a wider range of output values, and vice versa for bright input values. When $c = \gamma = 1$ the gamma transformation reduces to the identity transformation. The modified gamma transformation proposed in this work used $c = 1$ and $\gamma = 0.7$, by means a piece-wise function in which the final value of each pixel depends on a certain threshold value θ. In particular, the function mapping the initial value of a pixel $r(x, y)$ into its final value $s(x, y)$ is defined in Equation (15). The value of θ, used for the piece wise gamma selection, is the one suggested by the Otsu's thresholding method [30]. Figure 4 shows the results of each pre-processing step applied on two (one per row) DCE-MRI breast lesions.

$$s(x,y) = \begin{cases} r(x,y) & if \quad r(x,y) \leq \theta) \\ r(x,y)^{\gamma} & if \quad r(x,y) > \theta) \end{cases}. \tag{15}$$

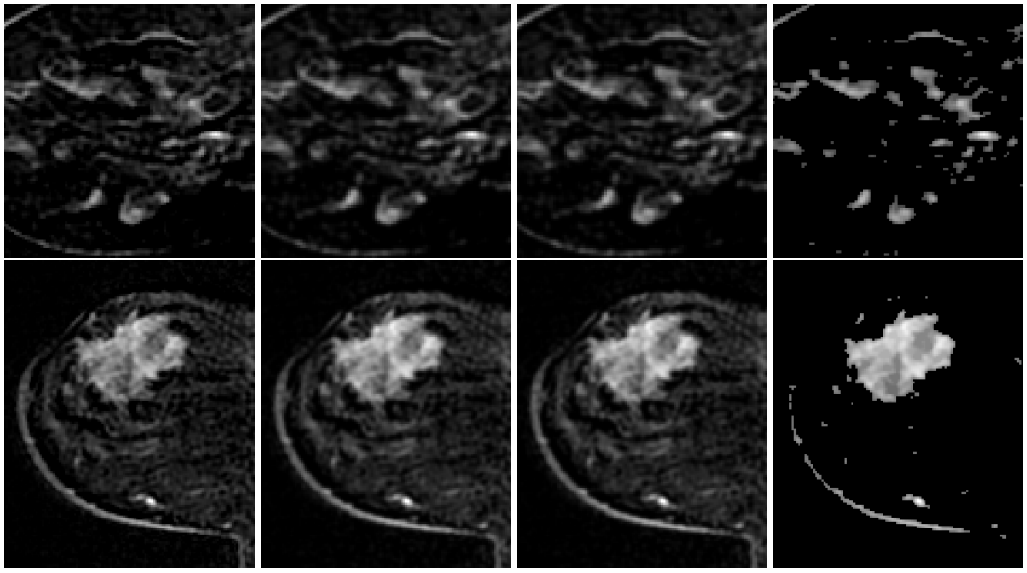

Figure 4. Examples of pre-processing steps on two breast lesions (one per row): (**1st column**) DCE-MRI breast image (after cropping); (**2nd column**) Gaussian smoothing result; (**3rd column**) centering result; and (**4th column**) gamma correction result. With respect to the position in the DCE-MRI sequence, all images have been rotated 90° clockwise, in order to improve the graphic representation and allow all the images relating to a specific lesion to be displayed in a single line.

3.2.3. Segmentation

After pre-processing, the segmentation involves the partitioning of a digital image into multiple sets of pixels, according to the specific clinical purpose. Nowadays, segmentation of digital images still remains one of the most challenging topics in image processing: in fact, medical image segmentation is often still performed manually, via time-consuming and operator-dependent procedures.

As indicated in Section 3.2.1, to start the segmentation process, it is necessary to select the ROI containing the lesion in the central image. The central slice—which is generally the one with the largest tumor section—is determined automatically after the operator sets the range of the initial and final slices containing the tumor. Starting from the selected ROI in the central slice, the ROIs in the adjacent slices are determined starting from the centroid of the segmented lesion in the previous slide. By doing so, the ROIs (set automatically) in the adjacent slices are centered (or almost) with the lesion section they contain. After the segmentation, a check is made to verify that the lesion does not touch the ROI boundary. In fact—considering that the ROI size is set equal to that of the ROI in the previous slice—it could happen that the selected ROI is too small and, consequently, the lesion was cropped. If so, the ROI is enlarged in order to fully contain the lesion section and repeated segmentation.

The segmentation is sequential: starting from the central slice, all the upper slices are processed first and then the lower ones. All the segmentation steps are the same for both the central slice and the other slices. The only difference lies in the different definition of the ratios—defined in Equations (18) and (19)—used in the post-processing steps, differentiated with the goal of optimizing the segmentation result.

SMRG Setting

There are several parameters controlling the SMRG behavior.

First of all, the splitting predicate P, a logical predicate fundamental to achieve satisfactory outputs. In order to identify breast lesions, homogeneity criteria are defined in terms of the mean value μ of each quad-regions. Here, the logical predicate in Equation (16) was used:

$$\mathsf{P} := \begin{cases} \mathsf{TRUE} & \text{if} \quad (\mu > 0.1) \wedge (\mu < 0.6) \\ \mathsf{FALSE} & \text{otherwise} \end{cases} . \tag{16}$$

The minimum block dimension $\rho_{\min}$ sets the minimum quad-region size beyond which no further splitting is carried out: the best results are found with minimum block dimensions of 4×4 pixels, because small regions are detected too. After the splitting phase, the final partition will contain adjacent regions with identical properties: this drawback can be addressed by merging only adjacent regions whose combined pixels satisfy the predicate P.

Finally, a stopping rule interrupts the growing procedure if no more pixels match the membership criterion. For the aim of this work, a reasonable stopping rule uses a similarity criterion between the candidate pixel to be incorporated and the pixels already belonging to the region. The criterion for the stopping rule is defined in terms of absolute distance between the regional mean of each quad-region and the threshold provided by the Otsu's method [30] on the original cropped image, both of them calculated without the contribution of null pixels.

Clustering Setting

Here, the optimal setting used for k-means, FCM and sFCM clustering algorithms is reported, in terms of the following parameters: *(i)* number of clusters; *(ii)* maximum number of iterations; *(iii)* minimum improvement in objective function between two consecutive iterations; *(iv)* exponent for the fuzzy partition matrix (only for fuzzy approaches); *(v)* $< p, q >$ parameters (only for fuzzy approaches).

Regarding the number of clusters, the lower the partition fuzziness, the better the segmentation result: the best clustering is achieved when V_{pe} is minimal. If the partition entropy for $C = 2$ is lower than the one obtained for $C = 3$, the lesion does not include any necrotic region: if so, pixels belonging the brightest cluster are turned to 1 and the others to 0. Otherwise, the two brightest clusters are fused together and their pixels are turned to 1, allowing for the inclusion of necrotic cores into the preliminary mask. Considering that, sometimes breast lesions can include an inner necrotic region that appears darker (i.e., hypo-intense) with respect to the rest of the lesion. In those cases where a binary clustering ($C = 2$) does not allow for properly detecting a necrosis, the number of clusters might be increased ($C = 3$). The choice of the optimal number of clusters to be used in k-means, FCM and sFCM is automatically set by evaluating the partition entropy for both cases $C = 2$ and $C = 3$. The partition entropy V_{pe} is a cluster validity function defined as in Equation (17):

$$V_{pe} = \frac{-\sum_{j=1}^{N}\sum_{i=1}^{C} u_{ij} \log u_{ij}^2}{N}. \tag{17}$$

The $\langle p, q \rangle$ parameters, dealing with FCM and sFCM clustering, control the relative importance of both functions defined in Equations (10) and (11), respectively. In particular, with $\langle p, q \rangle = \langle 1, 1 \rangle$, we sFCM is equivalent to the traditional FCM. As a matter of fact, to properly weight spatial information, the values $\langle p, q \rangle >= \langle 1, 2 \rangle$ were used. The choice was guided by the analysis and results obtained in [18], tackling a similar problem on DCE-MRI images. All these parameters are reported in Table 2.

At the end of the segmentation of each slice, the algorithm also verifies whether or not the contour of the detected lesion touches the boundaries of the initial rectangular bounding-box: if so, the shape of the lesion results in a cut version of the original one. For this reason, when the lesion intersects the boundaries of the original $m \times n$ rectangular crop, the algorithm shifts and expands the bounding-box itself until the boundaries of both the lesion and the crop do not intersect each other.

Table 2. Parameters setting for k-means, FCM and sFCM clustering algorithms.

Unsupervised Algorithm	Number of Clusters	Maximum Number of Iterations	Minimum Tolerance	Exponent of Fuzzy Partition Matrix	$\langle p, q \rangle$ Parameters
k-means	2 or 3	100	N.A.	N.A.	N.A.
FCM	2 or 3	100	1×10^{-5}	2.0	<1,1>
sFCM	2 or 3	100	1×10^{-2}	2.0	<1,2>

3.2.4. Post-Processing

The result of the segmentation process is a binary image in which a label is assigned to each pixel: 0 if the pixel does not belong to the mask, and 1 otherwise. The aim of the post-processing phase is to properly choose, among the connected-components resulting from the previous step, the only one representing the lesion. The post-processing itself consists of several sub-steps allowing for removing connected-components that do not meet specific morphological criteria (Figure 5).

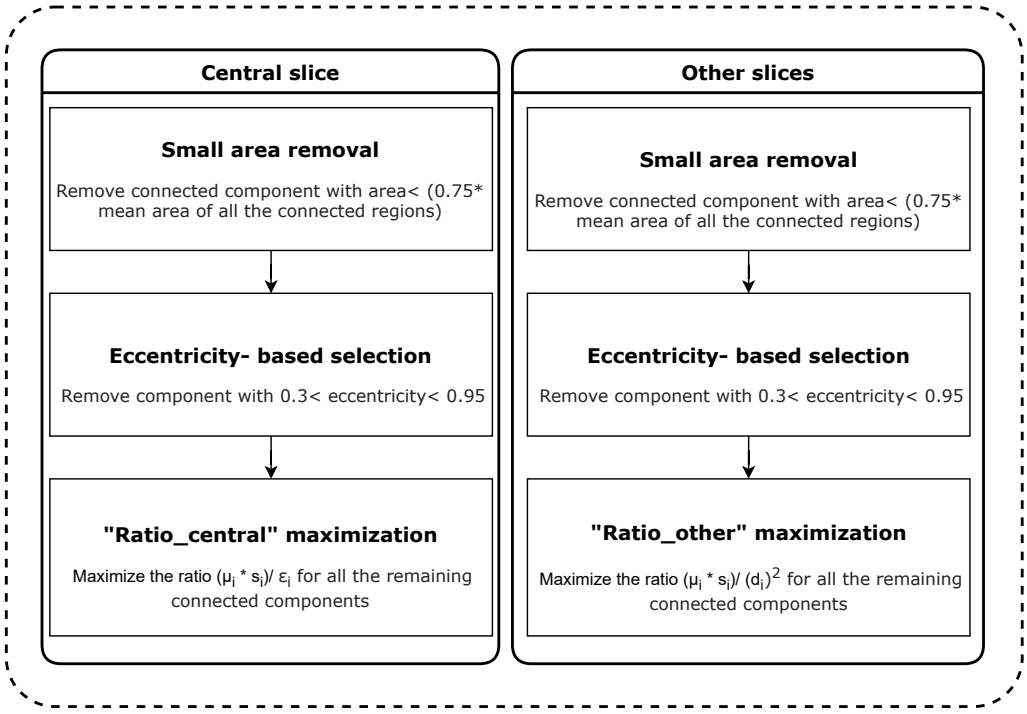

Figure 5. Diagram of the post-processing steps for both central slice and other slices.

The first post-processing phase allows us to extract all the connected-components from the raw binary mask. For each connected-component, a list of morphological features (i.e., extreme points, area, centroid, eccentricity, mean intensity and solidity) are computed that the algorithm is going to exploit in the following steps. In order to remove spurious connected-components with an area that is too small to be considered as a lesion, an area-based selection is performed: all the connected-components with an area smaller than $0.75 \times$ meanArea of all the connected-components. In such a way, spurious regions with a very small area are removed, reducing the number of valid candidates to be analyzed in the following steps.

Clinical evidence allows us to state that breast lesions generally grows in an isotropic way preserving a pseudo-spherical appearance [39]. As a consequence, their particular shape can be described in terms of eccentricity, which is defined as the ratio of the distance between the *foci* of an ellipse and its major axis length. Eccentricity values lay in the range $[0, 1]$: the extreme values represent degenerate cases identifying a circle and a line segment, respectively. The aim of this post-processing step is to delete from the list of the lesion candidates the regions with eccentricity lower than 0.3 and higher than 0.9 (these threshold values were determined experimentally). Eccentricity values close to 1 relate to very elongated lesions, while eccentricity values close to 0 relate to almost perfectly round shapes. In the case of breast lesions, even if overall there are rounded lesions, the lesion has a 'lobed' trend for which the final value of the eccentricity never takes values below 0.2–0.4. Using 0.3 as eccentricity lower limit in the post-processing phase, allow us to eliminate those (almost perfectly) circular connected components with pixel values similar to the lesion (and which are therefore incorrectly selected), but which are instead part of the background. As a consequence of their tendency to be round-like shaped, breast lesions exhibit an high solidity, which is defined as the ratio between the region area and the (including the region) convex polygon area. Considering this morphological information, the lesion ROI identification can be turned into an optimization problem with the ultimate goal of maximizing ratioCentral, defined in Equation (18):

$$\text{ratioCentral}_i = \frac{\mu_i s_i}{\epsilon_i}, \tag{18}$$

where μ_i is the mean intensity, s_i is the solidity and ϵ_i denote the eccentricity of the i-th connected-component, respectively.

The post-processing in the other slices differs from the central slice. In this case, the parameter to maximize ratioOther is defined according to Equation (19):

$$\text{ratioOther}_i = \frac{\mu_i s_i}{d_i^2}, \tag{19}$$

where d_i^2 represents the square of the distance between the centroid of the lesion region in the central slice and the centroids of all the connected-components in the current slice. In order to reduce false positives identification, ratioOther maximization searches for the connected-component with the centroid that is the closest to the one of the central slice mask. Recalling that breast lesions usually preserve a pseudo-spherical appearance, as the slices move away from the central one, the cross section identifying the lesion is reduced. In Equation (19) the replacement of ϵ_i with d_i^2 aims to avoid misclassification of spurious regions in slices different from the central one.

Figure 6 shows the results of each post-processing step applied to a segmentation mask obtained by means of sFCM clustering. Unfortunately, breast lesions might be characterized by different scenario—in terms of uniformity, contrast and well-defined boundaries—which a simple thresholding cannot properly manage. Figure 7 shows two segmentation results that allow us to appreciate the lesion non-homogeneity that only clustering approaches can properly manage, thus maximizing the result accuracy.

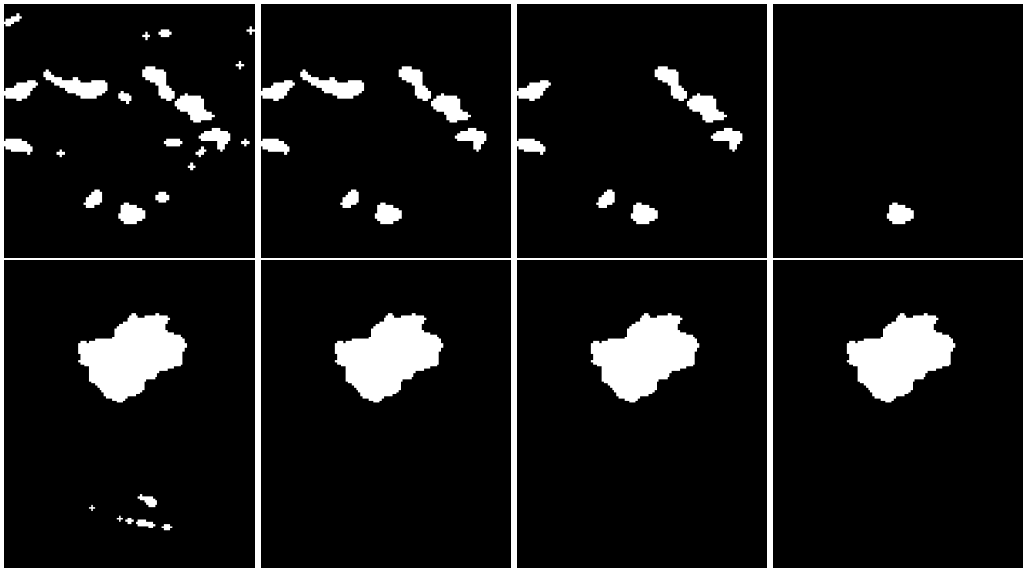

Figure 6. Examples of post-processing steps on two breast lesions (one per row): (**1st column**) binary mask obtained after sFCM clustering; (**2nd column**) mask after the small-area removal; (**3rd column**) mask after both the small-area removal and the eccentricity-based selection; and (**4th column**) mask after ratio criteria, defined in Equations (18) and (19). With respect to the position in the DCE-MRI sequence, all images have been rotated 90° clockwise, in order to improve the graphical representation and allow all the images, related to aspecific lesions, to be displayed in a single line.

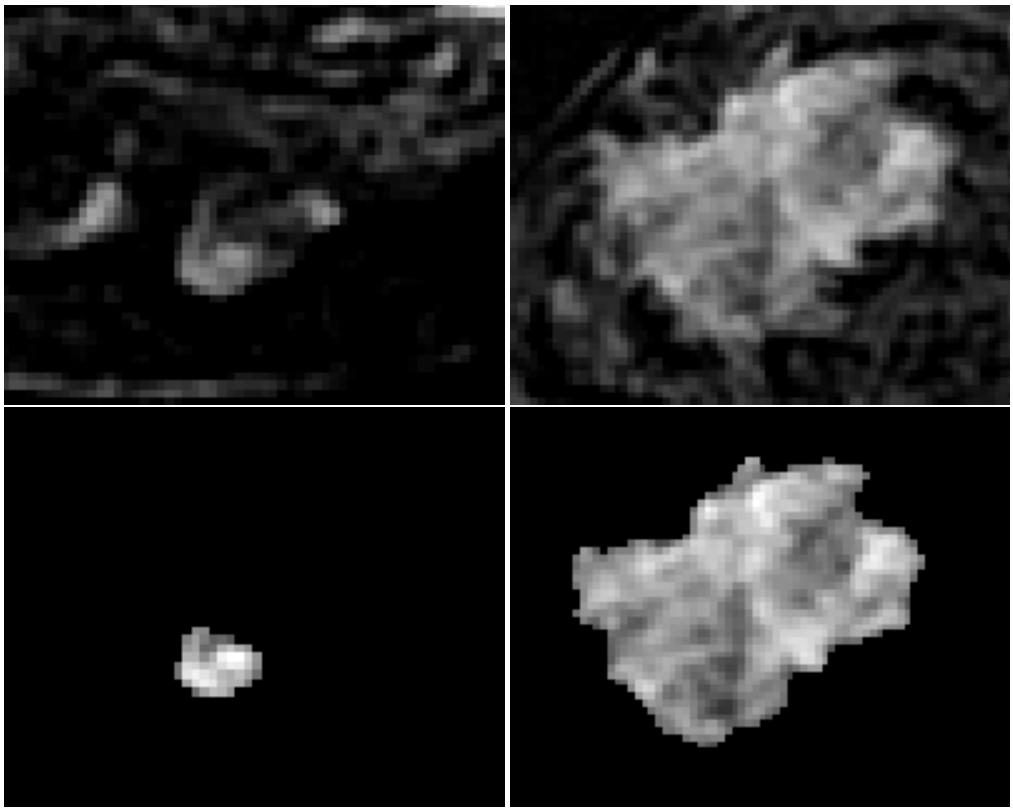

Figure 7. Two examples of segmentation obtained by sFCM: (**1st row**) the original DCE-MRI images; (**2nd row**) the corresponding segmentation. Compared to the images shown in Figures 4 and 6, a 2× zoom factor was used, that allows us to better appreciate the lesion non-homogeneity that only clustering approaches can properly manage, thus maximizing the final result accuracy.

4. Segmentation Performance Evaluation

Several evaluation measures are computed to quantify segmentation performances by comparing masks—obtained from analyzed unsupervised methods—and the ground-truth—provided by a radiologist, with more than 5-year experience on breast MRI, in consensus with a consultant breast radiologist (with more than 30-year experience on breast imaging). To obtain an accurate and detailed quantification, both area-based and distance-based metrics were used. The reason behind this choice is that area-based metrics strongly depend on region size and are not always able to evaluate the precision of a segmentation approach. On the other hand, distance-based metrics take into account the distance between the boundaries of the two segmentations to be compared, ignoring the actual volume difference between the two masks.

4.1. Spatial Area-Based Metrics

Spatial area-based metrics compare the semi-automatic segmented regions with the manually segmented ones ($\mathcal{R}_A$ and $\mathcal{R}_T$, respectively) by calculating the overlapping percentage of area between the two masks obtained from the segmentation of the image $\mathcal{I}$. Recalling some basics on statistical decision theory measures, the regions containing 'true positives' (TP), 'false positive' (FP), 'false negatives' (FN), and 'true negatives' (TN) are defined as:

$$\begin{aligned}\mathcal{R}_{\mathrm{TP}} &= \mathcal{R}_A \cap \mathcal{R}_T \\ \mathcal{R}_{\mathrm{FP}} &= \mathcal{R}_A - \mathcal{R}_T \\ \mathcal{R}_{\mathrm{FN}} &= \mathcal{R}_T - \mathcal{R}_A \\ \mathcal{R}_{\mathrm{TN}} &= \mathcal{I} - \mathcal{R}_T - \mathcal{R}_A\end{aligned}$$

When validating the segmentation results, the two most used area-based metrics are the Dice Index (DI) and the Jaccard Index (JI), defined in Equations (20) and (21), respectively. DI and JI are used to describe how much similar the manual (ground-truth) and the semi-automatic segmentations are: the greater they are, the higher is the overlapping percentage between the two masks.

$$\mathrm{DI} = \frac{2 \cdot \mathcal{R}_{\mathrm{TP}}}{\mathcal{R}_A + \mathcal{R}_T} \tag{20}$$

$$\mathrm{JI} = \frac{\mathcal{R}_A \cap \mathcal{R}_T}{\mathcal{R}_A \cup \mathcal{R}_T} \tag{21}$$

Sensitivity and specificity—defined in Equations (22) and (23)—represent the portion of positive pixels (foreground) and negative pixels (background) correctly detected by a segmentation method with respect to the ground-truth, respectively.

$$\text{Sensitivity} = \frac{\mathcal{R}_{\mathrm{TP}}}{\mathcal{R}_{\mathrm{TP}} + \mathcal{R}_{\mathrm{FN}}} \tag{22}$$

$$\text{Specificity} = \frac{\mathcal{R}_{\mathrm{TN}}}{\mathcal{R}_{\mathrm{TN}} + \mathcal{R}_{\mathrm{FP}}} \tag{23}$$

False Positive Ratio (FPR) and False Negative Ratio (FNR)—defined in Equations (24) and (25)—denote the presence of false positives and false negative compared to the reference region, respectively.

$$\mathrm{FPR} = \frac{\mathcal{R}_{\mathrm{FP}}}{\mathcal{R}_{\mathrm{FP}} + \mathcal{R}_{\mathrm{TN}}}, \tag{24}$$

$$\mathrm{FNR} = \frac{\mathcal{R}_{\mathrm{FN}}}{\mathcal{R}_{\mathrm{FN}} + \mathcal{R}_{\mathrm{TP}}}. \tag{25}$$

4.2. Spatial Distance-Based Metrics

Area-based metrics are susceptible to differences between the positions of segmented regions and strongly dependent on their own size. To take into account the spatial position of the pixels, it is necessary to quantify the distance between the boundaries computed by the semi-automatic methods and the ground-truth delineated by the expert. Let $A = \{a_i : i = 1, 2, \ldots, K\}$ be the set of vertices belonging to the semi-automatic mask and $T = \{t_j : i = 1, 2, \ldots, N\}$ the set of vertices belonging to the ground-truth, the distance between the i-th pixel in A and the set T is defined as:

$$d(a_i, T) = \min_{j \in \{1,2,\ldots,N\}} ||a_i - t_j||_2, \tag{26}$$

where $||a_i - t_j||_2$ denotes the Euclidean distance between two points.

Many metrics can be defined in order to quantify the similarity/dissimilarity between two segmentations.

The Mean Absolute Distance (MAD)—defined in Equation (27)—quantifies the average error in the segmentation process. The Maximum Distance (MaxD)—defined in Equation (28)—measures the maximum difference between the two ROI boundaries. The Hausdorff Distance (HD) between the point sets A and T—defined in Equation (29)—measures the maximal distance from a point in the first set to a nearest point in the other one.

$$\mathrm{MAD} = \frac{1}{K} \sum_{i=1}^{K} d(a_i, T) \tag{27}$$

$$\mathrm{MaxD} = \max_{i \in 1,2,\ldots,K} \{d(a_i, T)\} \tag{28}$$

$$\mathrm{HD} = \max \{h(T, A), h(A, T)\}, \tag{29}$$

where $h(T, A) = \max_{t \in T} \{\min_{a \in A} \{d(t, a)\}\}$ is the so-called 'directed Hausdorff Distance'.

It is important to point out that all the measured distances are expressed in pixels: in this way, they result will be independent from the spatial resolution among different MRI datasets (i.e., pixel spacing).

5. Experimental Findings

5.1. Area-Based Metrics Segmentation Results

Area-based metrics obtained by each segmentation algorithm are shown in Tables 3 and 4—expressed as mean ± standard deviation. As easily appreciable, the results showed that the fuzzy framework offered by FCM and sFCM reflects the intrinsic uncertainty that characterizes medical images, allowing us to achieve better segmentation results compared to the hard clustering performed by k-means. Furthermore, it is worth to note that spatial constraints taken into account by sFCM contribute to reduce the standard deviation of the final result, thus ensuring higher reliability with respect to SMRG and k-means. On the other hand, the large value in standard deviation indicates that SMRG has a high variability of the results that affects its reliability. The boxplots in Figures 8 and 9 summarize the obtained results.

Regarding sensitivity, the results showed that the FCM-based approach offered better performance in terms of both mean value and standard deviation; on the contrary, the hard partition offered by k-means did not provide satisfying results. The specificity values showed that the clustering-based approaches obtained better performances compared to SMRG. In particular, sFCM and k-means achieved slightly better results than FCM.

Altogether, the results in Tables 3 and 4 showed that segmentation approaches based on soft-clustering techniques achieved better performance compared to SMRG and k-means. In fact, by explicitly exploiting the fuzziness, both FCM and sFCM better handled the intrinsic uncertainty and the natural variability of medical images. On the other hand, the crisp k-means allowed us to reach satisfying results in specificity values, but without granting good performance in sensitivity.

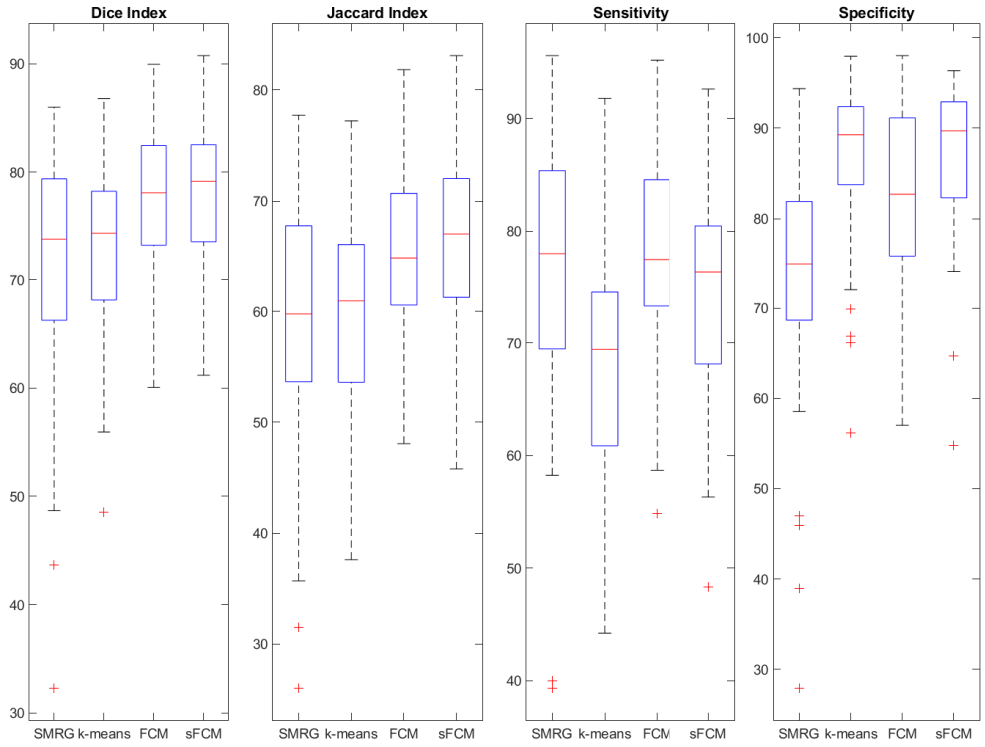

Figure 8. Boxplots of the area-based metrics achieved by the considered unsupervised segmentation approaches. From left to right: Dice Index, Jaccard Index, Sensitivity and Specificity. The lower and upper bounds of each box represent the first and the third quartiles of the metric distribution, respectively. The median is represented by a red line, while outliers are displayed as red crosses.

Table 3. Area-based metrics achieved by the considered unsupervised segmentation approaches: the results are expressed as average value ± standard deviation.

Method	DI	JI	Sensitivity	Specificity
SMRG	71.82 ± 10.83	58.98 ± 11.32	76.47 ± 12.41	73.65 ± 13.50
k-means	72.65 ± 8.26	59.87 ± 9.20	67.23 ± 10.78	86.97 ± 8.93
FCM	77.48 ± 6.77	65.22 ± 8.13	**77.84 ± 8.72**	81.94 ± 10.41
sFCM	**78.23 ± 6.50**	**65.90 ± 8.14**	74.69 ± 9.39	**87.10 ± 8.24**

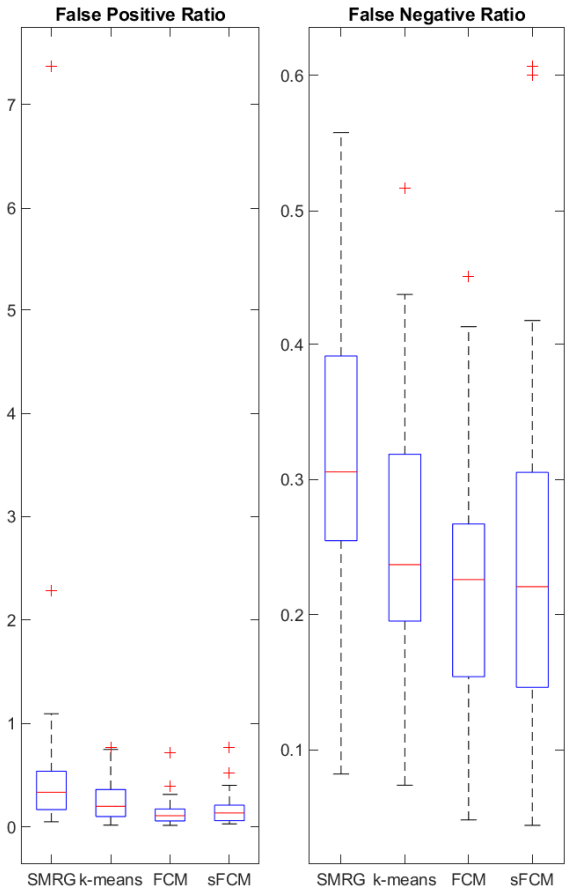

Figure 9. False Positive Ratio and False Negative Ratio boxplots. The lower and upper bounds of each box represent the first and the third quartiles of the metric distribution, respectively. The median is represented by a red line, while outliers are displayed as red crosses.

Table 4. False Positive and False Negative Ratios achieved by the considered unsupervised segmentation approaches: the results are expressed as average value ± standard deviation.

Method	False Positive Ratio	False Negative Ratio
SMRG	0.55 ± 1.05	0.33 ± 0.11
k-means	0.25 ± 0.20	0.25 ± 0.09
FCM	0.16 ± 0.13	0.23 ± 0.12
sFCM	**0.14 ± 0.12**	**0.22 ± 0.09**

Figure 10 shows some segmentation examples focusing on scenarios where the results are not particularly satisfactory in terms of FPs and FNs. In general, it is possible to note that SMRG and k-means have a greater tendency to leave out parts of the lesion (FN)—especially, when these are not uniform—and to include areas outside the lesion (FP), while the techniques based on clustering, and in particular the sFCM guarantees a better lesion detection.

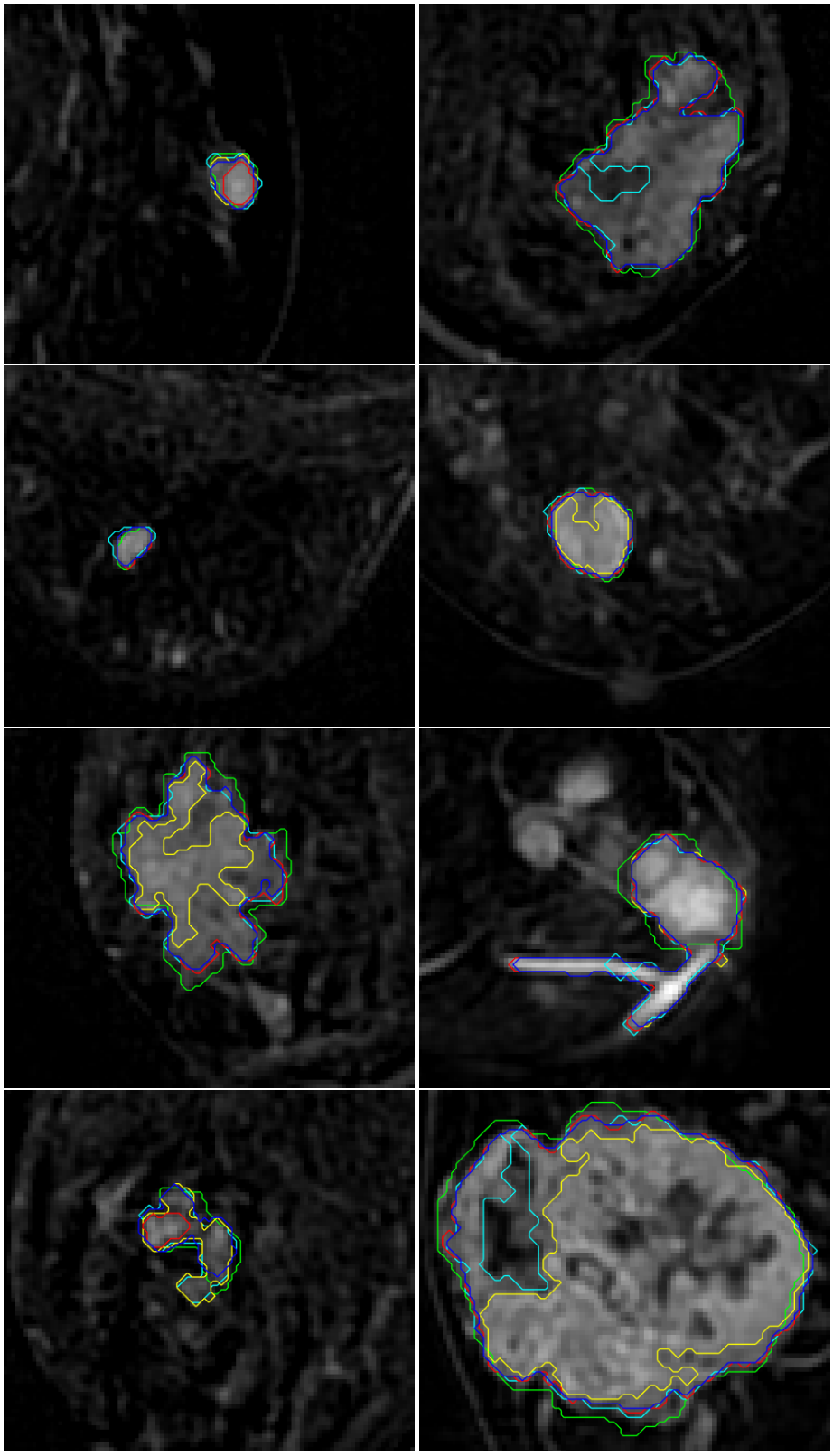

Figure 10. Examples of interesting segmentation results showing FPs and FNs. In particular, the examples compare results yielded by the four investigated unsupervised method—namely, SMRG (cyan), k-means (yellow), FCM (red), sFCM (blue)—against the ground-truth (green).

5.2. Spatial Distance-Based Metrics Segmentation Results

The need for using both area- and distance-based metrics comes out in considering that area-based metrics do not take in account pixels' spatial distribution. This leads to the need to quantify the distance between the boundaries computed by the semi-automatic methods and the ground-truth. The boxplots in Figure 11 summarize the obtained results.

In terms of HD, all the presented methods share similar characteristics with a mean value of ≈2.2 and a standard deviation of ≈0.43. This means that the boundaries of the semi-automatic masks are quite close to the manually traced ones.

The MAD metric is quite similar across the examined techniques, except for the SMRG, which shows the higher mean value and standard deviation. Clustering-based approaches, on the other hand, result more stable and precise, offering a better performance on the whole dataset.

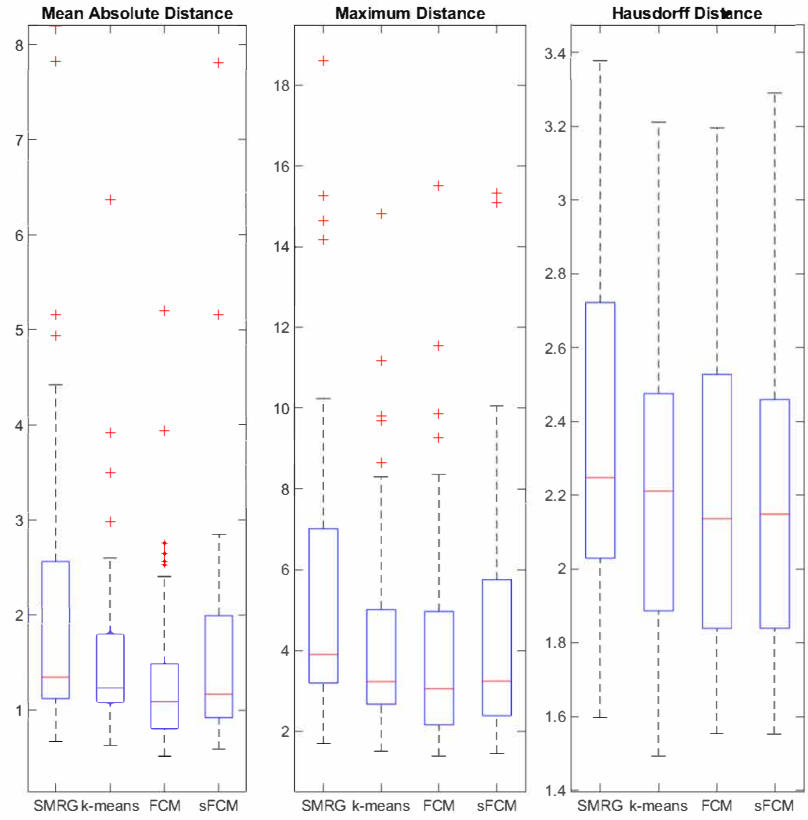

Figure 11. Boxplots of the spatial distance-based metrics achieved by the considered unsupervised segmentation approaches. From left to right: Mean Absolute Distance, Maximum Distance and Hausdorff Distance. The lower and upper bounds of each box represent the first and the third quartiles of the metric distribution, respectively. The median is represented by a red line, while outliers are displayed as red crosses.

Table 5 shows distance-based metrics obtained using the considered unsupervised segmentation approaches: lower distance values indicate better segmentation results. Observing the general trend, just a small deviation between the segmentations of the proposed methods and those of the experienced radiologist can be denoted. Furthermore, the achieved spatial distance-based indices are consistent with area-based metrics, confirming that clustering-based segmentation approaches allow us to reach better results with respect to SMRG.

Table 5. Spatial distance-based metrics achieved by the proposed segmentation approaches: the results are expressed as average value ± standard deviation.

Method	MAD	MaxD	HD
SMRG	2.00 ± 1.41	5.63 ± 3.81	2.32 ± 0.44
k-means	1.58 ± 0.99	4.22 ± 2.73	2.24 ± 0.42
FCM	1.58 ± 1.22	4.46 ± 3.12	**2.21 ± 0.43**
sFCM	**1.37 ± 0.90**	**4.04 ± 2.87**	2.21 ± 0.44

As shown at the top of Figure 12, FCM (with $p = 6.268 \times 10^{-6}$ and $p = 2.670 \times 10^{-7}$) and sFCM (with $p = 8.566 \times 10^{-5}$ and $p = 1.855 \times 10^{-7}$) clustering methods achieved significantly higher DI values compared to SMRG and k-means, respectively.

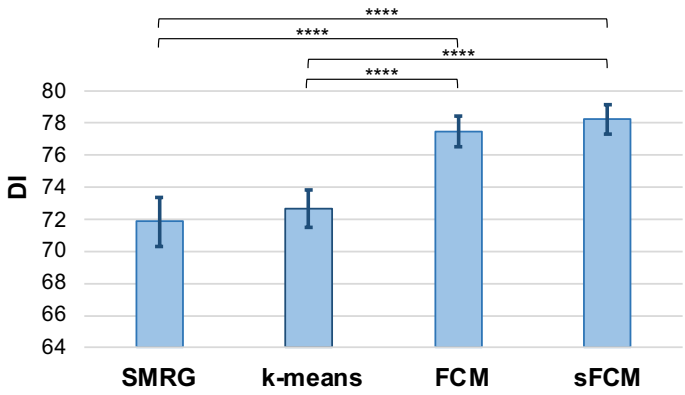

Figure 12. DI values achieved by the investigated traditional classic approaches. The bar graph and error bars denote the average value and the standard deviation DI values, respectively. The *p*-values, obtained from the statistical validation procedure, are shown at the top of the bars as brackets. The Wilcoxon rank-sum test for pairwise result comparison was used with the alternative hypothesis that the samples do not have equal medians of DI values. A significance level of $\alpha = 0.05$ with a correction using the Bonferroni–Holm method. Notation: **** $p < 0.0001$.

To statistically validate the obtained results, the two-sided Wilcoxon signed rank test [50] on paired DI results was performed with the null hypothesis that the samples come from continuous distributions with equal medians (considering a significance level of 0.05). Obtained *p*-values are shown in Table 6. With more details, this test on paired results was used to statistically compare the distributions of the DI values achieved by two competing methods and identify significant differences. The *p*-values were corrected by the Bonferroni–Holm method [51] for multiple comparisons.

The values reported in Tables 3–5 represent the average value ± standard deviation over all 50 breast masses. All the values obtained on each breast mass are reported in the Supplementary Materials.

Table 6. *p*-Values obtained from the statistical validation procedure. The Wilcoxon rank-sum test for pairwise result comparison was used with the alternative hypothesis that the samples do not have equal medians of DI. A significance level of $\alpha = 0.05$ with the Bonferroni-Holm correction for multiple comparisons was used. **Boldface** indicates that the null hypothesis can be rejected.

	SMRG	k-Means	FCM
k-means	0.973		
FCM	$\mathbf{6.268 \times 10^{-6}}$	$\mathbf{2.670 \times 10^{-7}}$	
sFCM	$\mathbf{8.566 \times 10^{-5}}$	$\mathbf{1.855 \times 10^{-7}}$	0.423

5.3. Processing Times

In order to evaluate the processing times, all segmentations were performed on the entire dataset, by calculating the average elapsed time and the corresponding standard deviation for each of the four investigated algorithms over all the analyzed images, obtaining the following values: SMRG: 2.78 ± 2.79 s; k-means: 1.35 ± 0.78 s; FCM: 1.65 ± 0.88 s; sFCM: 1.76 ± 0.83 s. These processing times were measured using the Matlab R2019b IDE (by means of the `tic` and `toc` stopwatch timer functions) running on a Windows 10 Pro general-purpose PC equipped with an Intel I7-3630QM@2.40 GHz CPU and 8 GB RAM.

As expected, SMRG had the longest processing time due to the two-stage approach and also considering the iterative processes employed during the Split-and-Merge and Region Growing executions. Interestingly, the introduction of the fuzzy logic is negligible compared to the crisp k-means, as well as the integration of the spatial constraints into the sFCM algorithm does not require a remarkable computational overhead in addition to the standard FCM clustering.

Overall, these results demonstrate the clinical feasibility of the investigated classic unsupervised methods also in terms of both computational resources and processing times, by considering that supervised CNN-based approaches for segmentation require a training phase and then an inference phase typically performed on Graphics Processing Units [18].

5.4. Difficult Cases

As previously highlighted, medical images are characterized by an intrinsic variability in which boundaries or anatomical details may be not well defined. Furthermore, noise corrupts digital images, thus affecting certain features within the original image. MRI suffers from various kinds of noise and artefacts because of the nature of the signal detection and spatial encoding [37,38]. For instance, hardware-induced errors are often caused by the complicated acquisition scheme depending on radiofrequency coils, while thermal noise can derive from transmission lines, receiver circuits and polarization magnetic field B_0 drift during the scan acquisitions. In addition, natural body motion (e.g., respiratory and cardiac motion) can degrade the image quality too. As a consequence, even after a proper pre-processing step, it is common to deal with low-contrast, noisy images. The proper lesion segmentation in this kind of images is not a trivial task and user interaction would be required to produce accurate results.

5.4.1. Case with Low Contrast-Enhanced Mass

In some cases, MR images are difficult to segment because the lesion itself does not appear brighter with respect to the muscle and adipose surrounding tissue. This kind of images exhibit a narrow histogram located typically toward the middle of the intensity scale, implying a washed-out grey look through the whole image. As a consequence, the meaningful partition of the original image results in a very difficult task leading to imprecise results. The opposite is true for the histogram of a high-contrast image, which covers a wide range of the intensity scale and has a pixel distribution not too far from uniform. The effect is an image that shows a great deal of gray-level detail and has a high dynamic range. As a matter of fact, a segmentation process on this kind of images will produce satisfying results with a high reproducibility. As easily appreciable in Figure 13, MR images offer a very difficult scenario where the lesion boundaries are not clearly distinguishable from the rest of the image. The manual segmentation (green contour) cannot properly segment the whole lesion because of the strong uncertainty due to the low contrast characterizing the whole image. On the other hand, the FCM segmentation (red contour) correctly identifies the lesion but, because of the low percentage of overlapping area with the manual mask, it does not ensure satisfying results.

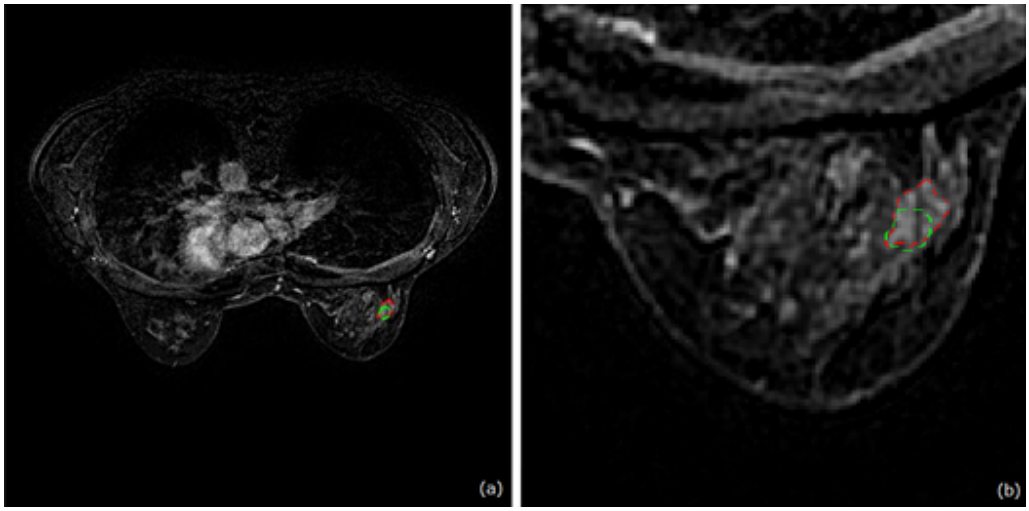

Figure 13. Case with a low-contrast lesion: (**a**) entire axial DCE-MRI slice; (**b**) crop of the slice shown in (**a**). Comparison between manual ground-truth (green) and automatic segmentation (red).

5.4.2. Case with Blurred Boundary Mass

From Figure 14, it is possible to appreciate how blur severely compromises the contrast of the original image, making lesion segmentation a very challenging task. As mentioned above, because of the strong uncertainty in boundaries delineation, the manual segmentation (green contour) identifies a very simple and smooth shape into which the lesion is included. Of course, this kind of strategy allows the identification of lesion's location, but also includes into the ROI a lot of false positives. On the other hand, the semi-automatic mask (red contour) identifies a more precise region avoiding the misclassification of FP pixels. Even in this case, because of the imperfect area, overlap area-based metrics will not yield a high score.

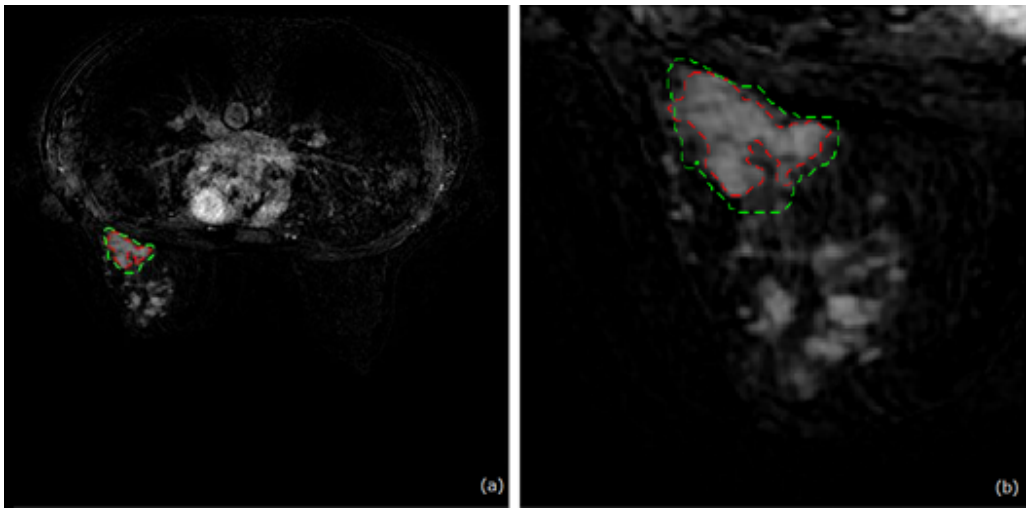

Figure 14. Case with a blurred boundary lesion: (**a**) entire axial DCE-MRI slice; (**b**) crop of the slice shown in (**a**). Comparison between manual ground-truth (green) and automated segmentation (red).

5.4.3. Case with Irregular Mass

Opposed to what is reported in the literature [39], breast lesions may sometimes exhibit irregular shapes and borders with internal divisions. In these cases, breast masses could cause difficulties during the manual segmentation process: in fact, as the lesion contour becomes more irregular, the manual tracing of the ROI becomes more challenging. As a consequence, the manual segmentation of breast lesions with unusual shapes does not always match the effective lesion contour. Figure 15 exhibits one of this cases with unusual

elongated breast lesion. The manual segmentation (green contour) completely cut the left portion in the upper part of the lesion which is, instead, properly included in the semi-automatic boundary (red contour). As a consequence, computer-assisted segmentation process results in a more precise mask that properly reproduces the narrowed-shape of lesion.

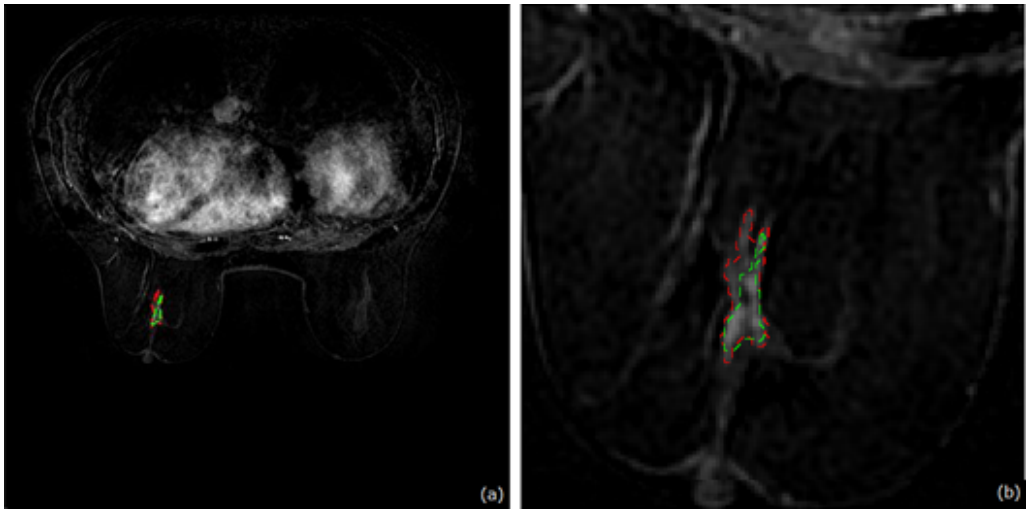

Figure 15. Case with an irregular lesion: (**a**) entire axial DCE-MRI slice; (**b**) crop of the slice shown in (**a**). Comparison between manual ground-truth (green) and automated segmentation (red).

6. Discussion and Conclusions

The main objective of this work was to offer a detailed analysis of well-established classical unsupervised segmentation techniques by carefully comparing them in a real clinical application. Breast cancer is the most common cause of cancer death in women worldwide [52,53] and the second most common cancer overall [54]. Fortunately, science evolution has led to the development of medical imaging techniques, which are used to detect abnormalities in breast parenchyma. Among imaging techniques, multiparametric MRI plays a crucial role and it is widely used in clinical applications, due to its high resolution images and the ability to precisely differentiate soft tissues.

As matter of fact, as a case study, the contrast-enhancing mass delineation in breast DCE-MRI was addressed by means of four popular unsupervised segmentation methods, namely: SMRG, k-means, FCM, and sFCM. Although they represent well-known approaches in the literature, they are still widely used in clinical tools. Starting from the basic versions of these approaches, during the initial analysis, we identified the shortcomings of each of them, developing and implementing improved versions, when possible.

Nowadays, deep learning techniques represent the state-of-the-art, allowing us to achieve high performance and accurate lesion segmentation for datasets of thousands of patients [14]. Deep learning approaches for image segmentation are generally supervised techniques that require a considerable computation times and a large amount of data for training [15]. In fact, these data must be representative of all the possible scenarios in which the deep neural network could operate, not always available in small- or medium-sized hospitals, and therefore not clinically feasible. Moreover, it should be noted that the authors of [18] showed that semi-automatic approached based on classic unsupervised techniques obtained results comparable or superior to the deep CNNs (namely, SegNet and U-Net). Therefore, this study was focused on classic pattern recognition approaches with the goal of providing an in-depth analysis.

It is important to point out that, even considering all the disadvantageous aspects related to manual segmentation, the contribution of an expert radiologist still remains essential at least to validate the results obtained by means a computer-assisted approach. In fact, clinicians rely on computational approaches with interpretable results [41]. From this point of view, the classical unsupervised approaches—such as those analyzed in this

work—provide this important advantage. This aspect is even more critical in deep learning architectures, where CNNs are generally adopted as 'black-box', thus making it difficult to offer a physical interpretation to the features encoded in all the intermediate CNN layers [46].

The obtained experimental results, in terms of area- and distance-based metrics, encourage the use of unsupervised pattern recognition techniques in medical image segmentation. In particular, consistently with [55], clustering-based segmentation approaches achieved better performance compared to the SMRG, the only 'thresholding-based' approach considered. As a consequence, crisp segmentation techniques—such as k-means and SMRG—are not well-suited for medical images that are characterized by an uncertain/variability (sometimes related to the noise), yielding inaccurate boundaries and not well-defined details. Both FCM and sFCM clustering techniques—implementing fuzzy modeling that provides an intrinsic flexibility—significantly achieved the best performance. In fact, on area-based metrics, they obtained DI = 78.23% ± 6.50 (sFCM), JI = 65.90% ± 8.14 (sFCM), sensitivity = 77.84% ± 8.72 (FCM), and specificity = 87.10% ± 8.24 (sFCM), FPR = 0.14 ± 0.12 (sFCM), and FNR = 0.22 ± 0.09 (sFCM). On distance-based metrics, they obtained MAD = 1.37 ± 0.90 (sFCM), MaxD = 4.04 ± 2.87 (sFCM), and HD = 2.21 ± 0.43 (FCM).

A second segmentation of the same radiologist or the segmentation of a different radiologist would have certainly allowed us to quantify the inter-/intra-operator variability of the results. Nevertheless, as already observed in [56], the mean DI was 0.81 (range 0.19–0.96). The mean DI is higher for the 'easy tumors' compared to the 'challenging tumors' (0.83 *vs.* 0.75, respectively, $p < 0.001$). The mean DI for each observer combination separately, for all tumors, ranged between 0.78 and 0.83, where the segmentations of the breast radiologist and the medical student showed the highest overlap. These results confirm that the performance achieved by the best performing methods are in line with the inter-observer agreement, also in terms of metrics variability according to the lesion types.

As further developments, we plan to investigate innovative improvements to further improve the performance with fuzzy clustering, by using (*i*) more sophisticated membership functions, and (*ii*) more advanced pre- and post-processing steps. Moreover, investigating and comparing the latest machine learning techniques, such as Generative Adversarial Networks (GANs), for unsupervised detection and segmentation [57,58] would be relevant with a sufficient amount of data for training and test. Finally, the implementation of multiparametric or multimodal approaches [59], by using different types of co-registered medical images—i.e., Diffusion Weighted Imaging (DWI) and Positron Emission Tomography (PET)/MRI—probably would allow us to improve the detection performance [60,61].

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/app12010162/s1: Table S1: Dice Index (DI) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S2: Jaccard Index (JI) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S3: Sensitivity values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S4: Specificity values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S5: False Positive Ratio (FPR) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S6: False Negative Ratio (FNR) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S7: Mean Absolute Distance (MAD) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S8: Maximum

Distance (MaxD) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported; Table S9: Hausdorff Distance (HD) values obtained by the four investigated unsupervised methods on each of the 50 segmented breast masses on DCE-MRI. In the last row the mean value ± the standard deviation is reported.

Author Contributions: Conceptualization, C.M. and A.R.; methodology, C.M. A.R. and L.R.; software, A.R.; validation, C.M. and L.R.; formal analysis, C.M. and L.R.; investigation, A.R.; resources, C.M., L.R. and I.D.; data curation, C.M. and I.D.; writing—original draft preparation, C.M., A.R. and L.R.; writing—review and editing, F.M., T.V.B., F.B. and G.R.; visualization, C.M. and L.R.; supervision, F.M., T.V.B., F.B. and G.R.; project administration, T.V.B., F.B. and G.R.; funding acquisition, L.R. and G.R. All authors have read and agreed to the published version of the manuscript.

Funding: This study has received funding by GeSeTon project, funded by Italian MISE Grant No. 489 of 21 February 2018. This study has also been partially supported by The Mark Foundation for Cancer Research and Cancer Research UK Cambridge Centre [C9685/A25177] and by the Royal Society for the International Exchanges 2020 Cost Share with the Italian CNR (project No. IEC/R2/202313). Additional support was also provided by the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre [BRC-1215-20014]. 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.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of "Azienda Ospedaliera Universitaria Policlinico P.Giaccone" of Palermo, Italy (protocol code n.1/2020-15/01/2020).

Informed Consent Statement: Retrospective data collection was approved by the Ethics Committee. The requirement for evidence of informed consent was waived because of the retrospective nature of our study.

Data Availability Statement: The data presented in this study may be available on reasonable request from the corresponding author. The data are not publicly available due to ethical and privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Badr, E. Images in Space and Time: Real Big Data in Healthcare. *ACM Comput. Surv.* **2021**, *54*, 113. [CrossRef]
2. Duncan, J.S.; Ayache, N. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 85–106. [CrossRef]
3. Ghavami, N.; Hu, Y.; Gibson, E.; Bonmati, E.; Emberton, M.; Moore, C.M.; Barratt, D.C. Automatic segmentation of prostate MRI using convolutional neural networks: Investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration. *Med. Image Anal.* **2019**, *58*, 101558. [CrossRef]
4. Lee, B.; Yamanakkanavar, N.; Choi, J.Y. Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture. *PLoS ONE* **2020**, *15*, e0236493. [CrossRef]
5. Pasero, E.; Castagneri, C. Application of an automatic ulcer segmentation algorithm. In Proceedings of the International Forum on Research and Technologies for Society and Industry (RTSI), Modena, Italy, 11–13 September 2017; pp. 1–4. [CrossRef]
6. Yankeelov, T.E.; Mankoff, D.A.; Schwartz, L.H.; Lieberman, F.S.; Buatti, J.M.; Mountz, J.M.; Erickson, B.J.; Fennessy, F.M.; Huang, W.; Kalpathy-Cramer, J.; et al. Quantitative imaging in cancer clinical trials. *Clin. Cancer Res.* **2016**, *22*, 284–290. [CrossRef] [PubMed]
7. Aerts, H.J.; Velazquez, E.R.; Leijenaar, R.T.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [CrossRef] [PubMed]
8. Lambin, P.; Leijenaar, R.T.; Deist, T.M.; Peerlings, J.; de Jong, E.E.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.; Even, A.J.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749. [CrossRef]
9. Krupinski, E.A. Current perspectives in medical image perception. *Atten. Percept. Psychophys.* **2010**, *72*, 1205–1217. [CrossRef] [PubMed]
10. Rundo, L.; Pirrone, R.; Vitabile, S.; Sala, E.; Gambino, O. Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine. *J. Biomed. Inf.* **2020**, *108*, 103479. [CrossRef]
11. Lee, N.Y.; Lu, J.J. *Target Volume Delineation and Field Setup: A Practical Guide for Conformal and Intensity-Modulated Radiation Therapy*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]

12. Hamamci, A.; Kucuk, N.; Karaman, K.; Engin, K.; Unal, G. Tumor-Cut: Segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications. *IEEE Trans. Med. Imaging* **2012**, *31*, 790–804. [CrossRef]
13. Rundo, L.; Militello, C.; Russo, G.; Vitabile, S.; Gilardi, M.C.; Mauri, G. GTVcut for neuro-radiosurgery treatment planning: An MRI brain cancer seeded image segmentation method based on a cellular automata model. *Nat. Comput.* **2018**, *17*, 521–536. [CrossRef]
14. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
15. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef]
16. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [CrossRef] [PubMed]
17. Ravì, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.Z. Deep learning for health informatics. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 4–21. [CrossRef] [PubMed]
18. Militello, C.; Rundo, L.; Dimarco, M.; Orlando, A.; Conti, V.; Woitek, R.; D'Angelo, I.; Bartolotta, T.V.; Russo, G. Semi-automated and interactive segmentation of contrast-enhancing masses on breast DCE-MRI using spatial fuzzy clustering. *Biomed. Signal Process. Control.* **2022**, *71*, 103113. [CrossRef]
19. Acharya, R.; Wasserman, R.; Stevens, J.; Hinojosa, C. Biomedical imaging modalities: A tutorial. *Comput. Med. Imaging Graph.* **1995**, *19*, 3–25. [CrossRef]
20. Tirpude, N.N.; Welekar, R.R. Effect Of Global Thresholding On Tumor-Bearing Brain MRI Images. *Int. J. Eng. Comput. Sci.* **2013**, *2*, 728–731.
21. Militello, C.; Vitabile, S.; Rundo, L.; Russo, G.; Midiri, M.; Gilardi, M.C. A fully automatic 2D segmentation method for uterine fibroid in MRgFUS treatment evaluation. *Comput. Biol. Med.* **2015**, *62*, 277–292. [CrossRef] [PubMed]
22. Islam, M.R.; Imteaz, M.R.; Marium-E-Jannat. Detection and analysis of brain tumor from MRI by Integrated Thresholding and Morphological Process with Histogram based method. In Proceedings of the International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 8–9 February 2018; pp. 1–5.
23. Rundo, L.; Militello, C.; Vitabile, S.; Casarino, C.; Russo, G.; Midiri, M.; Gilardi, M.C. Combining Split-and-Merge and Multi-Seed Region Growing Algorithms for Uterine Fibroid Segmentation in MRgFUS Treatments. *Med. Biol. Eng. Comput.* **2016**, *54*, 1071–1084. [CrossRef]
24. Horowitz, S.L.; Pavlidis, T. Picture Segmentation by a Tree Traversal Algorithm. *J. ACM* **1976**, *23*, 368–388. [CrossRef]
25. Manousakas, I.N.; Undrill, P.E.; Cameron, G.G.; Redpath, T.W. Split-and-Merge Segmentation of Magnetic Resonance Medical Images: Performance Evaluation and Extension to Three Dimensions. *Comput. Biomed. Res.* **1998**, *31*, 393–412. [CrossRef]
26. Saad, N.M.; Abu-Bakar, S.A.R.; Muda, S.; Mokji, M. Automated segmentation of brain lesion based on diffusion-weighted MRI using a split and merge approach. In Proceedings of the IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, Malaysia, 30 November–2 December 2010; pp. 475–480.
27. Saad, N.M.; Abu-Bakar, S.A.R.; Muda, S.; Mokji, M.; Abdullah, A.R. Automated region growing for segmentation of brain lesion in diffusion-weighted MRI. In Proceedings of the International MultiConference of Engineers and Computer Scientists, IMECS 2012, Hong Kong, China, 14–16 March 2012; pp. 674–677.
28. Adams, R.; Bischof, L. Seeded region growing. *IEEE Trans. Pattern Anal.* **1994**, *16*, 641–647. [CrossRef]
29. Chang, Y.L.; Li, X. Adaptive image region-growing. *IEEE Trans. Image Process.* **1994**, *3*, 868–872. [CrossRef]
30. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1975**, *11*, 23–27. [CrossRef]
31. Joseph, R.P.; Senthil Singh, C.; Manikandan, M. Brain tumor MRI image segmentation and detection in image processing. *Int. J. Res. Eng. Technol.* **2014**, *3*, 1–5.
32. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.D.; Silverman, R.; Wu, A.Y. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [CrossRef]
33. Bezdek, J.C. Objective function clustering. In *Pattern Recognition with Fuzzy Objective Function Algorithms*, 1st ed.; Springer: Secaucus, NJ, USA, 1981; pp. 43–93. [CrossRef]
34. Li, Y.l.; Shen, Y. An automatic fuzzy c-means algorithm for image segmentation. *Soft Comput.* **2010**, *14*, 123–128. [CrossRef]
35. Militello, C.; Rundo, L.; Vitabile, S.; Russo, G.; Pisciotta, P.; Marletta, F.; Ippolito, M.; D'Arrigo, C.; Midiri, M.; Gilardi, M. Gamma Knife treatment planning: MR brain tumor segmentation and volume measurement based on unsupervised Fuzzy C-Means clustering. *Int. J. Imaging Syst. Technol.* **2015**, *25*, 213–225. [CrossRef]
36. Rundo, L.; Militello, C.; Russo, G.; Garufi, A.; Vitabile, S.; Gilardi, M.C.; Mauri, G. Automated prostate gland segmentation based on an unsupervised fuzzy c-means clustering technique using multispectral T1w and T2w MR imaging. *Information* **2017**, *8*, 49. [CrossRef]
37. Caponetti, L.; Castellano, G.; Corsini, V. MR brain image segmentation: A framework to compare different clustering techniques. *Information* **2017**, *8*, 138. [CrossRef]
38. Rundo, L.; Militello, C.; Tangherloni, A.; Russo, G.; Vitabile, S.; Gilardi, M.C.; Mauri, G. NeXt for neuro-radiosurgery: A fully automatic approach for necrosis extraction in brain tumor MRI using an unsupervised machine learning technique. *Int. J. Imaging Syst. Technol.* **2018**, *28*, 21–37. [CrossRef]
39. Feder, J.M.; de Paredes, E.S.; Hogge, J.P.; Wilken, J.J. Unusual breast lesions: Radiologic-pathologic correlation. *Radiographics* **2019**, *19*, S11–S26. [CrossRef] [PubMed]

40. Chuang, K.S.; Tzeng, H.L.; Chen, S.; Wu, J.; Chen, T.J. Fuzzy c-means clustering with spatial information for image segmentation. *Comput. Med. Imaging Graph.* **2006**, *30*, 9–15. [CrossRef] [PubMed]
41. Rundo, L.; Beer, L.; Ursprung, S.; Martin-Gonzalez, P.; Markowetz, F.; Brenton, J.D.; Crispin-Ortuzar, M.; Sala, E.; Woitek, R. Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering. *Comput. Biol. Med.* **2020**, *120*, 103751. [CrossRef] [PubMed]
42. Patil, S. Preprocessing To Be Considered For MR and CT Images Containing Tumors. *IOSR J. Electr. Electron. Eng.* **2012**, *1*, 55–57. [CrossRef]
43. Vasuki, P.; Kanimozhi, J.; Devi, M.B. A survey on image preprocessing techniques for diverse fields of medical imagery. In Proceedings of the 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), Tamilnadu, India, 27–28 April 2017; pp. 1–6.
44. Behrenbruch, C.; Petroudi, S.; Bond, S.; Declerck, J.; Leong, F.; Brady, J. Image filtering techniques for medical image post-processing: An overview. *Br. J. Radiol.* **2004**, *77*, S126—S132. [CrossRef] [PubMed]
45. Seeram, E.; Seeram, D. Image Postprocessing in Digital Radiology—A Primer for Technologists. *J. Med. Imaging Radiat. Sci.* **2008**, *39*, 23–41. [CrossRef]
46. Castiglioni, I.; Rundo, L.; Codari, M.; Di Leo, G.; Salvatore, C.; Interlenghi, M.; et al. AI applications to medical images: From machine learning to deep learning. *Phys. Med.* **2021**, *83*, 9–24. [CrossRef]
47. Motwani, M.C.; Gadiya, M.C.; Motwani, R.C.; Harris, F.C. Survey of image denoising techniques. *Glob. Signal Process. Expo Conf. (GSPX)* **2004**, *27*, 27–30.
48. Marrone, S.; Piantadosi, G.; Fusco, R.; Petrillo, A.; Sansone, M.; Sansone, C. Breast segmentation using Fuzzy C-Means and anatomical priors in DCE-MRI. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 1472–1477.
49. Piantadosi, G.; Fusco, R.; Petrillo, A.; Sansone, M.; Sansone, C. LBP-TOP for volume lesion classification in breast DCE-MRI. In *International Conference on Image Analysis and Processing*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; pp. 647–657. [CrossRef]
50. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1980**, *1*, 196–202. [CrossRef]
51. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
52. DeSantis, C.E.; Ma, J.; Gaudet, M.M.; Newman, L.A.; Miller, K.D.; Goding Sauer, A.; Jemal, A.; Siegel, R.L. Breast cancer statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 438–451. [CrossRef]
53. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2021. *CA Cancer J. Clin.* **2021**, *71*, 7–33. [CrossRef] [PubMed]
54. International Agency for Research on Cancer. The Global Cancer Observatory. 2020. Available online: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf (accessed on 28 July 2021).
55. Frackiewicz, M.; Koper, Z.; Palus, H.; Borys, D.; Psiuk-Maksymowicz, K. Breast lesion segmentation in DCE-MRI Imaging. In Proceedings of the 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 26–29 November 2018; pp. 308–313. [CrossRef]
56. Granzier, R.; Verbakel, N.; Ibrahim, A.; van Timmeren, J.; van Nijnatten, T.; Leijenaar, R.; Lobbes, M.; Smidt, M.; Woodruff, H. MRI-based radiomics in breast cancer: Feature robustness with respect to inter-observer segmentation variability. *Sci. Rep.* **2020**, *10*, 14163. [CrossRef] [PubMed]
57. Han, C.; Rundo, L.; Murao, K.; Noguchi, T.; Shimahara, Y.; Milacski, Z.Á.; Koshino, S.; Sala, E.; Nakayama, H.; Satoh, S. MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. *BMC Bioinf.* **2021**, *22*, 31. [CrossRef]
58. Wu, X.; Bi, L.; Fulham, M.; Kim, J. Unsupervised Positron Emission Tomography Tumor Segmentation via GAN based Adversarial Auto-Encoder. In Proceedings of the 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, 13–15 December 2020; pp. 448–453. [CrossRef]
59. Rundo, L.; Stefano, A.; Militello, C.; Russo, G.; Sabini, M.G.; D'Arrigo, C.; Marletta, F.; Ippolito, M.; Mauri, G.; Vitabile, S.; et al. A fully automatic approach for multimodal PET and MR image segmentation in Gamma Knife treatment planning. *Comput. Methods Programs Biomed.* **2017**, *144*, 77–96. [CrossRef]
60. Woitek, R.; McLean, M.A.; Gill, A.B.; Grist, J.T.; Provenzano, E.; Patterson, A.J.; Ursprung, S.; Torheim, T.; Zaccagna, F.; Locke, M.; et al. Hyperpolarized 13C MRI of tumor metabolism demonstrates early metabolic response to neoadjuvant chemotherapy in breast cancer. *Radiol. Imaging Cancer* **2020**, *2*, e200017. [CrossRef]
61. Carmona-Bozo, J.C.; Manavaki, R.; Woitek, R.; Torheim, T.; Baxter, G.C.; Caracò, C.; Provenzano, E.; Graves, M.J.; Fryer, T.D.; Patterson, A.J.; et al. Hypoxia and perfusion in breast cancer: Simultaneous assessment using PET/MR imaging. *Eur. Radiol.* **2021**, *31*, 333–344. [CrossRef]