



Enhancement of clustering techniques by coupling clustering tree and neural network: Application to brain tumour segmentation

Michael Osadebey¹  | Marius Pedersen¹ | Meeta Kalra² | Dag Waaler³ | Nizar Bouguila² 

¹Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway

²Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada

³Department of Health Sciences, Norwegian University of Science and Technology, Gjøvik, Norway

Correspondence

Michael Osadebey, Department of Computer Science, Norwegian University of Science and Technology, Teknologivegen 22, Gjøvik 2815, Norway.

Email: michael.osadebey@ntnu.no

Funding information

European Research Consortium for Informatics and Mathematics

Abstract

Currently, no classical clustering algorithm is efficient on its own. The predefined number of clusters required for their operation does not consistently produce satisfactory segmentation results. They exhibit cluster instability, are vulnerable to the local optimum trap, and are sensitive to noise and imaging artefacts. Most contributions designed to overcome these drawbacks incorporate prior knowledge such as cluster label information and statistic measures that demand minimal labelled training data. Although these approaches improve the segmentation accuracy, they tend to diminish the advantages of clustering algorithms over the supervised learning methods. This study proposes a shift from the use of a predefined number of clusters to a clustering tree-based method for performance enhancement of classical clustering algorithms. The proposed method is a three-stage algorithm. It begins with the extraction of low-level features from a clustering tree. Clustering trees are sets of labelled clusters of an image at multiple clustering resolutions. The second stage extracts high-level features by coupling the clustering tree to a single-layer feedforward neural network. The third stage is the classification stage, where the basic model of a neural network extracts the tumour from a high-level feature map. Because neither of the neural networks requires training, the proposed method is both fully unsupervised and fully automated and retains all its advantages over supervised methods. A performance evaluation using FLAIR MRI images of brain tumour patients from the BRATS2015 and BRATS2020 databases demonstrates significant performance enhancement over four classical clustering algorithms and two of the four proposed techniques were comparable to deep learning methods.

KEYWORDS

artificial neural network, clustering tree, expectation maximization (EM), fuzzy C-means (FCM), k-means (KM), Otsu-based multilevel thresholding (OTSU)

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Expert Systems* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Brain tumour is a neurological disease which manifest as the growth of abnormal structures within or close to the brain (Alibolandi et al., 2018). Recent reports shows that brain cancer is a significant burden on regional, national and global health-care systems (Patel et al., 2019). Gliomas, the common type of brain tumours, are classified by the world health organization (WHO) from grades 1 to 4 according to the tumour's growth potential and aggressiveness (Louis et al., 2021). Grades 1, 2 are the low grade gliomas (LGG), while grades 3, 4 are the high grade gliomas (HGG).

Magnetic resonance imaging (MRI) is widely used in diagnosis and follow-up examinations to classify a tumour and determine its shape, location and spatial extent (Lu et al., 2020). This information assist physicians to predict a patient survival and optimize treatment plan (Zarei Mahmoodabadi et al., 2013). The most attractive feature of MRI is the generation, from the same subject, of multispectral images which can be combined for accurate characterization and classification of brain tumour. Figure 1a–d are T1-weighted (T1-w), T1-weighted contrast-enhanced (T1c), T2-weighted (T2-w) and fluid attenuation inversion recovery (FLAIR) images, respectively (Menze et al., 2014). Although each MRI sequence provide complimentary information about the tumour and healthy brain structures, the contribution from the FLAIR sequence stands out from other sequences. The FLAIR signal enhances the visualization of tumour infiltration by suppressing the signals of water molecules so that the radiologist can conveniently distinguish the hyperintense tumour in the white matter region from the cerebrospinal fluid (CSF) in the ventricle (Regnery et al., 2019). The FLAIR sequence is the preferred imaging modality in radiotherapy planning to determine target volume for patient with HGG and to determine tumour extent and volume during surgery in LGG which generally do not reveal contrast enhancement (Edjlali et al., 2019; Soltaninejad et al., 2017).

Since routine brain tumour diagnosis for a single patient requires several hundreds of MRI slices, manual segmentation is cumbersome, prone to fatigue-induced errors with high intra-reader and inter-reader variability (Despotović et al., 2015). Furthermore, results from manual segmentation are not always reproducible and the segmentation accuracy is dependent on the radiologist experience. Therefore, automatic segmentation methods that incorporates general pathological information about tumours is necessary to facilitate image interpretation and improve diagnostic accuracy (Pereira et al., 2016). Contributions on automatic segmentation of brain tumour can be divided into four categories; threshold-based, region-based, model-based and pixel classification techniques (Gordillo et al., 2013). Threshold-based methods extract the tumour by comparing the intensity of the tumour region with one or more intensity thresholds (Li et al., 2016). In region-based techniques, the tumour is extracted by merging neighbourhood pixels with homogeneity properties based on predefined similarity criteria (Wong, 2005). Model-based methods extract the tumour region by using information built from a model of the brain structures that incorporates prior knowledge such as shape, location and orientation (Korez et al., 2015). This study will focus on clustering-based method of pixel classification techniques for the segmentation of brain

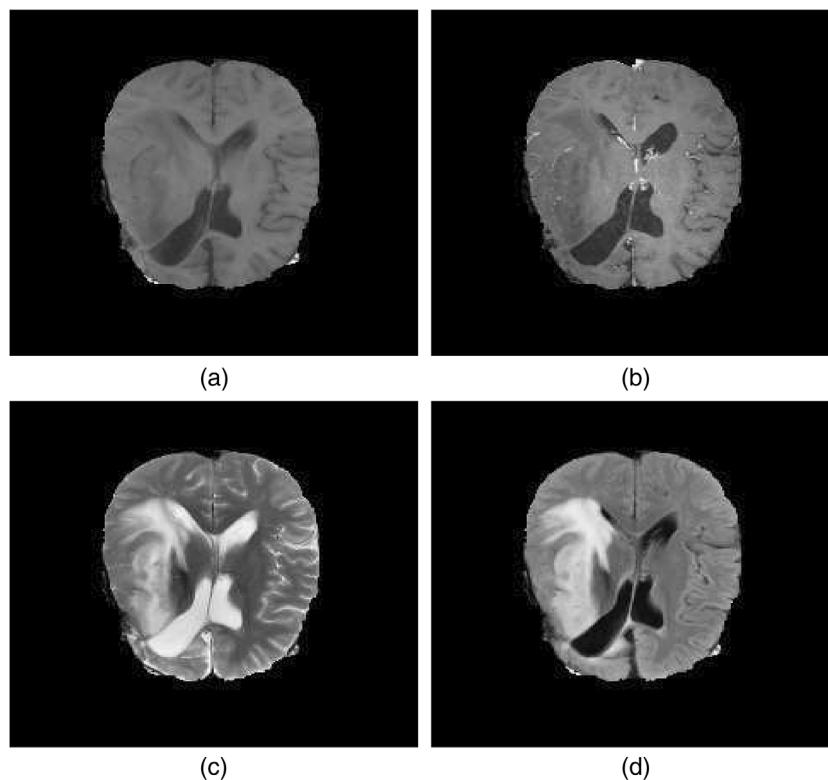


FIGURE 1 Four MRI sequence images from a patient (a) T1-weighted, (b) T1c-weighted, (c) T2-weighted, and (d) FLAIR

tumours. Readers who are interested in broader coverage of brain tumour segmentation can refer to Tiwari et al. (2020), Nazar et al. (2020), Kaur and Rani (2016), Işın et al. (2016), Wadhwa et al. (2019) and Nalepa et al. (2019).

1.1 | Outline of article

This article is organized as follows. Section 2 describes pixel classification techniques for the segmentation of brain tumour and highlights its drawbacks. Section 3 reviews some current contributions for enhancing the performance of clustering techniques. In Section 4, we propose a new approach to the segmentation of brain tumours in MRI images. Experiments to evaluate the proposed method and its results are describe in Section 5. Discussion is in Section 6. Section 7 concludes this article.

2 | PIXEL CLASSIFICATION TECHNIQUES

Pixel classification methods assume that pixels that describe the tumour region and other regions of the brain belong to a finite number of classes. Therefore, the pixels are analysed and grouped based on specific classification schemes (Gordillo et al., 2013). Two categories of pixel classification techniques are the supervised and unsupervised methods. Neural networks and clustering are the popular supervised and unsupervised pixel classification methods, respectively (Salem et al., 2009). Clustering is the process of grouping pixels into classes based on the minimization of an objective function. Since the past decade, deep neural networks are gaining wide popularity among researchers because of their outstanding performance in image segmentation. The convolutional neural network (CNN) is a type of deep neural network that is widely used for brain analysis. The UNet (Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., 2017) are two popular CNN-based deep neural networks for brain tumour segmentation. Both techniques apply encoding to extract contextual information from the input features and decoding to keep track of pixel locations in the encoding path. The decoding process in UNet is facilitated via skip connections, whereby the entire feature map in the encoder are transferred to the decoder whereas in SegNet only the pooling indices from the encoder are transferred to the decoder. For this reason, the SegNet uses less memory than UNet during training. There are two main challenges to achieve accurate segmentation with deep neural networks. First, the complex structure of both the CNN and the brain regions result in long training time and difficulty in the discrimination between tumour and none tumour regions. Second, government regulation on the protection of privacy limits the quantity and by extension the quality of medical data researchers can access for training a model. To reduce the training and computation time, recent contributions such as Badrinarayanan et al. (2017) include an attention mechanism in the CNN architecture to processes only relevant image regions rather than the whole brain. Other contributions propose the generative adversarial network (GAN) (Xue et al., 2018) as a component of the CNN for synthetic data generation to augment the limited availability of training data.

Other examples of techniques that incorporate neural networks is the contribution by (Lu et al., 2020) that applies batch normalization to fine-tune and improve the robustness of AlexNet (Krizhevsky et al., 2012), a pretrained deep convolutional neural network. Thereafter, several layers of the network are replaced by an extreme learning machine classifier (Ding et al., 2015), a training algorithm for single hidden layer feedforward neural network optimized using chaotic bat algorithm (Gandomi & Yang, 2014). Another neural network-based technique (Zhang et al., 2018) apply synthetic minority oversampling technique to balance the dataset. Pathological features extracted using wavelet packet Tsallis entropy were classified by extreme learning machine trained using Jaya algorithm. Below, we provide detailed description of clustering algorithms.

2.1 | Classical clustering algorithms

Four popular image clustering techniques are expectation maximization (EM) (Dempster et al., 1977), fuzzy c-means (FCM) (Bezdek et al., 1984), k-means (KM) (MacQueen et al., 1967) and Otsu-based multilevel thresholds (OTSU) (Huang et al., 2011; Merzban & Elbayoumi, 2019; Otsu, 1979). In this article, the four algorithms (in their original formulation) are referred to as classical clustering algorithms. Their implementation can be generalized into a four-step process: initialization, partitioning, updating, and iteration until convergence. The design and operational characteristics of the EM, FCM, KM and OTSU algorithms limit their performance in several applications. The cluster centres, which are selected randomly at the initialization stage, induce cluster instability. Cluster stability is the capability of a clustering algorithm to generate an identical partition of the data irrespective of the order in which the patterns are presented to the algorithm (Fahad et al., 2014). The consequences of cluster instability are inaccurate and inconsistent segmentation results (Limwattanapibool & Arch-int, 2017; Yao et al., 2013). The design of these algorithms does not incorporate spatial modelling, which considers the strong correlation among neighbouring pixels in MRI images. Therefore, they are sensitive to noise and intensity inhomogeneity (Zhang et al., 2019). Because these algorithms operate with unlabelled data, their optimal performance is dependent on overcoming the challenges encountered in the tuning of operational parameters. Presently, there is no widely

accepted automatic method for computing the number of clusters that optimizes the output of clustering algorithms. Present approaches to selecting a predefined number of clusters can be considered heuristic because they are based on user experience determined from prior knowledge of algorithm output (Bittmann & Gelbard, 2007). This encourages manual segmentation over automation because it compels the user to manually evaluate and repeat the entire segmentation operation if the output result is not satisfactory. Although the objective function incorporated into these algorithms is formulated to guarantee convergence to a local optimum, there is no guarantee that the convergence result will be accurate. This is because the algorithms can be trapped in a local optimum (Aggarwal & Singh, 2019; Qin et al., 2016).

The strong parametric assumption of the EM algorithm is not the best option in all applications as it may not represent the different characteristics of images that are observed in real-life scenarios (Bouguila et al., 2004). For example, in clinical environments, magnitude MRI images are widely used for diagnosing diseases. The pixels in magnitude MRI images follow the Rician distribution rather than the Gaussian distribution (which is assumed in the formulation of the EM algorithm) (Scrucca, 2021). In brain MRI images, normal tissues can be distinguished from each other and from tumours because they can be described by pixels within different bands of intensity levels. The intensity bandwidth that distinguishes the different tissues and tumours varies with the acquisition techniques, MRI systems, and image quality attributes. Hence, it is challenging to select a 'number of clusters' that will produce accurate and consistent segmentation results across slices in an MRI volume data and across different datasets. These drawbacks will be illustrated using BRATS2015 database images that are displayed in Figure 2. These images can be considered as the prototype of different characteristics of images in a clinical setting. The images in the first column (Figure 2a, e, i, and m) are FLAIR MRI images of patients with brain tumour disease. In a typical FLAIR MRI slice, the ventricle and grey matter structures are hypointense because they can be described by darker pixels. The white matter and tumours are described by brighter pixels although the tumour is hyperintense relative to the white matter. This prior knowledge of intensity bands associated with different brain structures will motivate a data scientist to set $K=3$ as the predefined number of clusters while using a classical clustering algorithm for segmentation. The data scientist anticipates that the algorithm will generate three clusters. One of the clusters will represent the ventricles and grey matter region, and the other two clusters will represent the white matter region and the tumour. The images in the second, third, and fourth columns of Figure 2 are the outputs of the k-means algorithm using predefined numbers of clusters, $k=3$, $k=4$ and $k=5$, respectively. Figure 2f will satisfy the expectation of the data scientist because the tumour was extracted accurately. The other images in the second column (Figure 2b, Figure 2j, and Figure 2n) will not do so because the tumour was not isolated from normal structures. In the third and fourth columns, where $k=4$ and $k=5$, respectively, a few pixels shifted to different clusters. For example, pixels that belonged to the tumour region in previous columns were assigned to different clusters in their present column. Based on this illustration, it is evident that a particular predefined number of clusters may not always produce accurate and consistent segmentation results across different slices and across different datasets.

2.2 | Artificial neural network

Artificial neural network (ANN) models how the human brain learns to recognize real-world objects (Bahra & Wiese, 2019). The ANN is one of several soft computing techniques in conjunction with, for example, support vector machine and genetic algorithm that can solve nonlinear problems, that cannot be described by a mathematical model (Zhang et al., 2018). Figure 3a shows the basic model of an artificial neural network. It is denoted as **BMNN** throughout this article. It consists of the inputs $\{X_1, X_2, \dots, X_n\}$ to a neuron q and the corresponding weighted connections $\{W_1, W_2, \dots, W_n\}$ between the n inputs and neuron. When each input X_i is propagated through the connections, its value is adjusted by the value of the connecting weight W_i . At the neuron, the information is processed by applying a transformation function f to the weighted sum of the inputs:

$$Y = f\left(b + \sum_i^n W_i X_i\right) \equiv f(R), \quad (1)$$

where b is a bias weight, $R = (b + \sum_i^n W_i X_i)$ is the input to the neuron, and Y is the output of the neuron. The activation function can be regarded as an electronic switch between the input to a neuron at a specific layer and its output to a neuron in another layer. The neuron is activated when the input R to the neuron is above a specified threshold. Otherwise, the neuron is considered to be deactivated.

Several activation functions are used in ANN. They include the binary step function, sigmoid function, and rectified linear unit. In this article, we have defined the binary step function $f(i, j)$ as a function that activates a neuron located at (i, j) in a 2D grid of dimension $M \times N$ only if the input R to the neuron satisfies specified threshold criteria:

$$f(R(i, j)) = \begin{cases} \bar{q}(i, j), & \tau_1 \leq R(i, j) \leq \tau_2, \\ 0, & \text{Otherwise} \end{cases}, \quad (2)$$

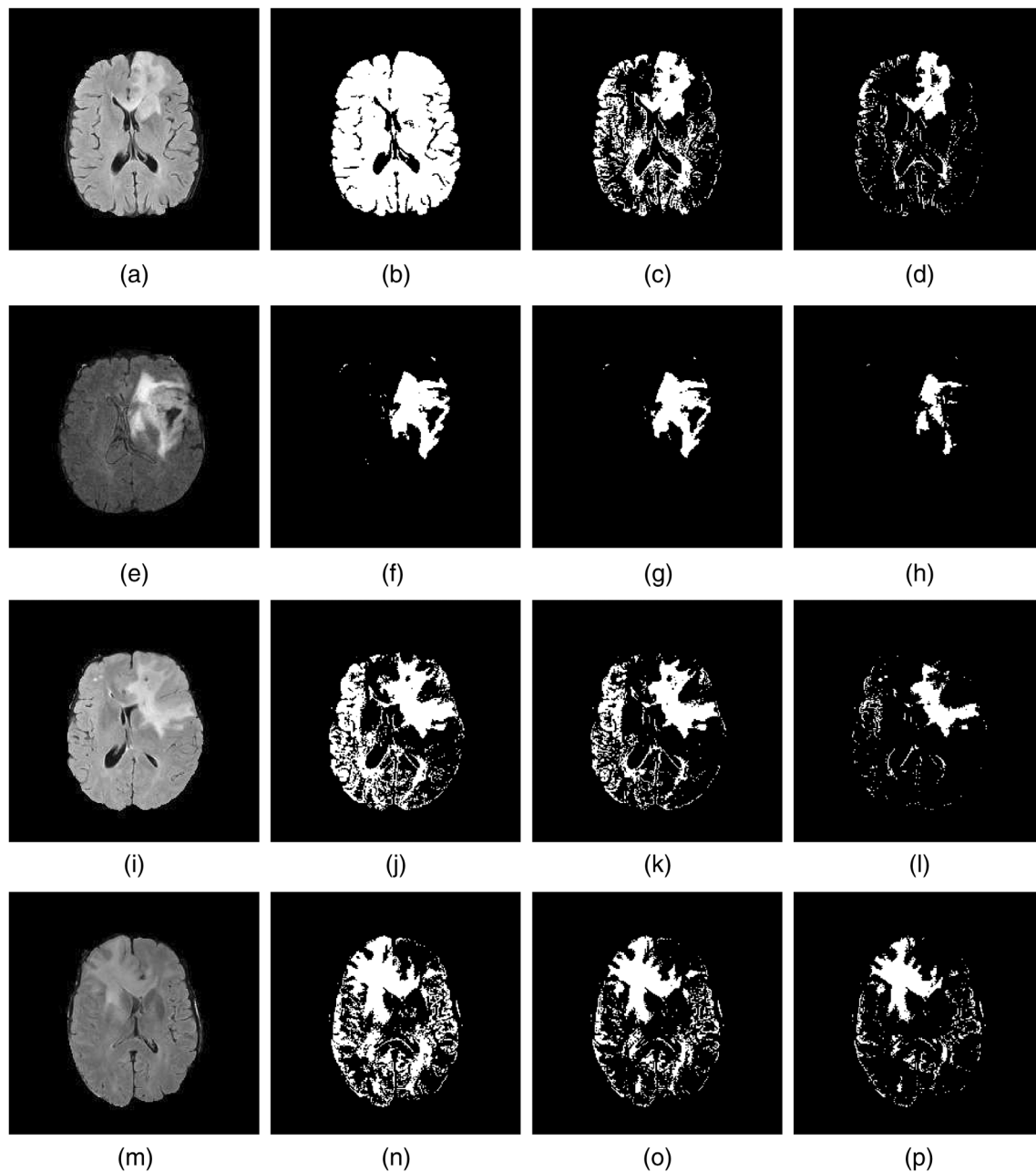


FIGURE 2 The first column, (a), (e), (i), and (m) are typical examples of FLAIR MRI images in a clinical setting. The second, third, and fourth columns are clusters extracted from corresponding images in the first column using k-means clustering algorithm with predefined numbers of clusters $k = 3$, $k = 4$, and $k = 5$, respectively

where $(\tau_1 \leq R(i, j) \leq \tau_2)$ is the specified threshold criteria. $\tilde{q}(i, j)$ is the transformed output from the activation function, which is sent to the next layer of neurons located at (i, j) in another 2D grid of identical dimension $M \times N$. The binary step function is unique for its simplicity and is considered the best option for executing unsupervised binary classification tasks. For supervised ANN, where backpropagation is required, the step function will be ineffective because its gradient is zero. The sigmoid function and the rectified linear unit are effective for an ANN that learns in a supervised manner. This is because they generally are differentiable, a prerequisite for the implementation of backpropagation.

Another basic ANN model is the single layer feedforward neural network. It is shown in Figure 3b and denoted as **SLFN** throughout this article. There is a neuron in the input layer, n neurons in the hidden layer, and a neuron in the output layer. For notational convenience, we denote the input and output neurons as X and Y , respectively. Each of the n branches that constitute the **SLFN** is equivalent to an input branch of the **BMNN** shown in Figure 3a. Therefore, the output from each activation function in the **SLFN** can be computed using the expression in Equation 1. Each activation function is fed to another transformation function g so that the output Y of the **SLFN** is

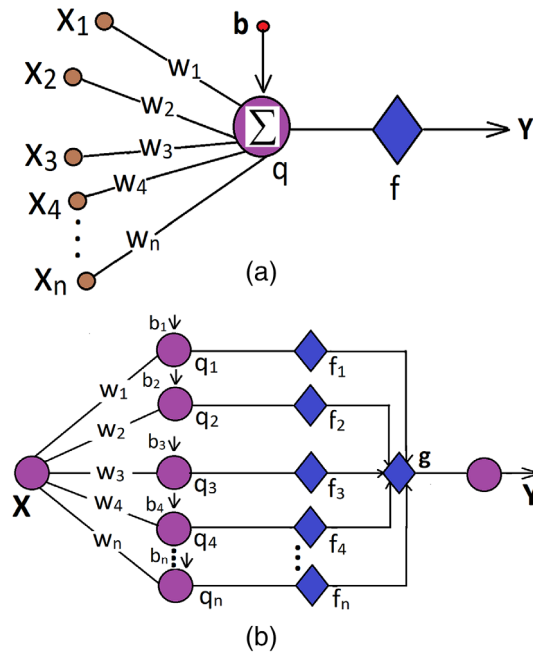


FIGURE 3 Schematic for implementing (a) basic model of an artificial neural network and (b) single-layer feedforward neural network

$$Y = g(f(b_i + W_i X_i)) \equiv g(f(R_i)). \quad (3)$$

2.3 | Clustering trees and scale-space multi-resolution analysis

Let $S(i, j) = \{s(1, 1), s(1, 2), \dots, s(M, N)\}$, where $s(i, j) \in \mathbb{R}^2$ represent the pixel intensity attributes of an image of dimensions $M \times N$. Furthermore, let us generate K duplicate copies of S . Each duplicate copy of S is assigned an index $k = \{1, 2, \dots, K\}$, where k is the number of clusters in which S_k will be partitioned. Now, let a classical clustering algorithm perform the labelling operation, which partitions each S_k into k clusters. Each cluster in S_k is represented by the matrix $C_p^k(i, j)$, where $1 \leq p \leq k$

$$C_p^k(i, j) = \begin{cases} p & \text{if } s(i, j) \text{ and } s(i', j') \text{ belong to same cluster, and } (i, j) \neq (i', j'), \\ 0 & \text{Otherwise.} \end{cases} \quad (4)$$

To clarify the above statement, the first duplicate image S_1 is partitioned into a cluster $C_1 = \{C_1^1(i, j)\}$. That is, the entire image is regarded as a cluster. The second duplicate image S_2 is partitioned into two clusters $C_2 = \{C_1^2(i, j), C_2^2(i, j)\}$, the third duplicate is partitioned into three clusters $C_3 = \{C_1^3(i, j), C_2^3(i, j), C_3^3(i, j)\}$, and the k th duplicate image S_k is partitioned into k clusters $C_k = \{C_1^k(i, j), C_2^k(i, j), \dots, C_k^k(i, j)\}$. The set of labelled clusters $C = \{C_1, C_2, \dots, C_k\}$ of an image at multiple clustering resolutions is referred to as clustering tree (Zappia & Oshlack, 2018). Throughout this article, clustering tree and clustering resolution are denoted as (CLTT) and κ , respectively. Each clustering resolution is a set of images derived from clustering at a predefined number of clusters $\{1 \leq k \leq K\}$. A clustering tree at clustering resolutions $\kappa = 4$ is shown in Figure 4a.

Clustering trees are conceptually similar to scale-space multi-resolution analysis. Scale-space multi-resolution analysis is a simultaneous representation of data at multiple scales (Lindeberg, 2007). Similarly, clustering tree is the simultaneous representation of clustered data at different clustering resolutions. A scale-space can reveal the constituent structures in a specified data at different scales. Clustering trees can provide insight into the heterogeneity of an image by revealing how pixels alter clusters with increasing clustering resolution (Von Luxburg, 2010). It is unfeasible to determine a priori the appropriate scales for describing structures within an image. Scale space theory is a reasonable approach to overcoming this limitation by describing structures at all scales to capture scale variations. It is challenging to determine a priori the number of clusters that describe the different characteristics of medical images in a clinical setting. Similar to scale space, clustering trees is a reasonable approach to overcoming this limitation because it considers clustering at all feasible clustering resolutions to capture more information on the constituent structures within the image. Because scale-space multi-resolution analysis considers all the feasible scales, it efficiently represents images at different levels of details (Xu et al., 2018). Because clustering trees consider all feasible clustering resolution, its efficiency is derived by identifying the structural information at different levels of segmentation accuracy (Jeub et al., 2018). Images generated at different clustering

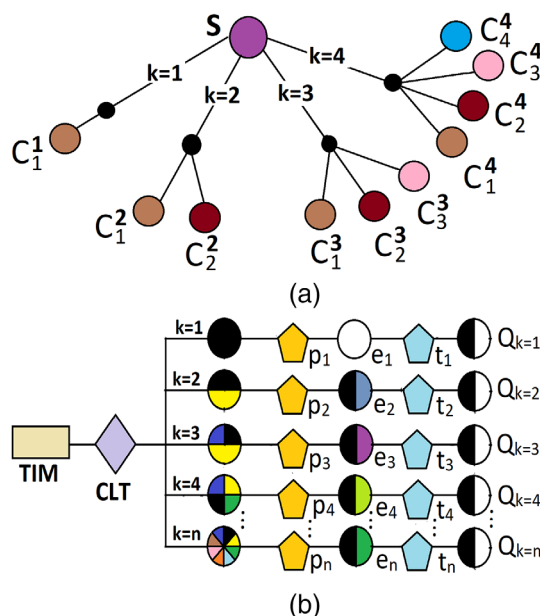


FIGURE 4 (a) A clustering tree clustering at resolution $\kappa = 4$ and (b) schematic of the implementation of a clustering tree at clustering resolution $\kappa = n$

resolutions potentially contain valuable local information, which can weaken the effect of noisy pixels. This enables it to extract different structures within the image accurately.

Notwithstanding the aforementioned similarities, clustering trees has two advantages over scale-space multi-resolution analysis. Unlike scale-space, the implementation of clustering trees does not require a kernel. Therefore, the edges and shape features are preserved in the images. In addition, there is no risk of either blurring or the introduction of extraneous features into the image. Furthermore, unlike scale-space multi-resolution analysis, the original image can be recovered accurately from the individual clusters in the clustering tree. Scale-space multi-resolution analysis has been applied to solve many image processing tasks such as noise removal (Pal et al., 2017), image matching (Lindeberg, 2015), and segmentation (Panigrahi et al., 2019; Zhuang et al., 2019). Applications using clustering trees are limited to tools for cluster analysis and cluster stability (Zappia & Oshlack, 2018). To our knowledge, there is presently no contribution describing clustering trees derived from the classical clustering algorithms, for image segmentation. We consider that researchers have not exhaustively explored the potential of clustering trees. Therefore, this article proposes a shift from a clustering technique based on a predefined number of clusters to a clustering tree-based segmentation method.

3 | RELATED WORK

Semi-supervised clustering is the popular technique for enhancing the performance of clustering algorithms. Semi-supervised methods utilize relatively marginal amounts of labelled data compared to the large volumes of unlabelled data utilized in supervised learning methods. The labelled data are used to simultaneously introduce a few attributes of supervised learning, tune algorithm parameters and minimize the drawbacks of the clustering algorithms. An example of semi-supervised clustering-based segmentation is the contribution by Portela et al. (2014). The operation of the algorithm begins with the random selection of relatively few brain MRI slices from several unlabelled MRI volume data. Thereafter, k-means clustering is applied to the selected data to generate clusters of regions-of-interest. The borders of the clustered regions-of-interest are refined and labelled by trained physicians. This is followed by the computation of statistical measures from the labelled regions-of-interest. The labelled information and statistical information are combined to form the initial parameters of a Gaussian mixture model-based clustering algorithm for segmenting a test image. Another semi-supervised clustering method was proposed by Saha et al. (2016). This method assumes that class labels are specified for 10% of the image data points, and the segmentation problem is posed as a multiobjective optimization problem based on internal and external cluster validity indices. The regions-of-interest were extracted using AMOSA, a simulated annealing-based multiobjective optimization technique proposed by Bandyopadhyay et al. (2008). Other examples of semi-supervised clustering-based methods include the contributions by Portela et al. (2014), Yang et al. (2020) and Wu and Zhang (2021). Although this approach enhances segmentation, quantifying the relatively marginal amount of labelled data is highly subjective. The quantification of labelled data may be dependent on the accuracy level of the algorithm,

and the risk of diminishing the advantages of fully unsupervised clustering over the supervised methods will be high. Furthermore, the generation of a reasonable amount of labelled data incurs additional cost because it requires hiring physicians with different levels of clinical experience.

Another technique for the enhancement of clustering algorithm is the incorporation of Markov random field (MRF) model within a two-stage clustering process. The first stage applied the fuzzy c-means algorithm for initial segmentation. In the second stage, the Markov random field models the local interactions between neighbouring pixels within the anatomical structures of a medical image where most pixels belong to an identical class as their neighbouring pixels. The output from the fuzzy algorithm and Markov random field model are cast within a Bayesian framework, and the segmented image is derived by maximizing the a posteriori probability of the segmentation given the image data. A contribution in this category, (Siyal & Yu, 2005) applies the classical fuzzy clustering to refine the grey level information in the original image and its associated multi-scale decomposition. Thereafter, the spatial constraints between neighbouring pixels in the image is modelled as a potential function in the Markov random field to reduce noise and enhance image quality. Another contribution (Saladi & Amutha Prabha, 2018) applies a modified fuzzy algorithm to estimate an initial segmentation parameters which are further applied to a Markov random field-based post processing. The contribution by Jafri et al. (2017) employs the expectation maximization algorithm to estimate hidden markov random field model of an image which also serves on the initial segmentation which is further refined using simple processing techniques. The incorporation of the MRF model improves the segmentation accuracy of clustering-based techniques by overcoming the sensitivity to noise and intensity inhomogeneity. However, Markov random field modelling is computationally intensive. Furthermore, the selection of optimal parameters for controlling the spatial interactions involves trade-off. For example, the segmented image becomes smooth with the loss of important structural information when the parameters of the spatial interactions are more than the optimal value. Moreover, the algorithm can be vulnerable to noise and intensity inhomogeneity when the selected parameters are less than optimal.

Combinations of multiple clustering techniques have been explored to enhance clustering algorithm. A four-stage algorithm proposed by Zhang et al. (2019) begins with denoising the original image using a wiener filter, followed by morphological operations to significantly eliminate structures outside the brain regions. The preliminary segmentation is generated by combining k-means clustering algorithm with Gaussian kernel-based fuzzy C-means algorithm in the third stage. The final segmentation is derived after combining another round of morphological operation with median filtering. The contribution by Shanker and Bhattacharya (2019) derive the initial cluster center using the classical k-means clustering. The initial segmentation was obtained by combining k-means and fuzzy clustering followed by the application of hierarchical centroid-based descriptor to extract the abnormal brain tissue. Hybrid methods combines attributes from the different clustering methods, however, the technique will still suffer from the trap of local minima and instability associated with the classical clustering algorithms.

Automatic determination of the predefined number of clusters have been explored in hybrid techniques to improve the performance of clustering algorithms. In Kaur and Sharma (2017), the original image is preprocessed to remove noise and enhance image quality, followed by histogram analysis which determines the appropriate number of clusters before k-means clustering for segmentation. In Pei et al. (2017), the number of clusters is automatically determined using the elbow rule (Mazurek & Mazurek, 2013) whereas a similar contribution (Ganesh et al., 2017) combines adaptive k-means and fuzzy c-means algorithms. The robustness of this approach is limited to images acquired from a specific scanner as the image features can easily be incorporated into the algorithm. More robust features need to be incorporated into the algorithm for images acquired from different scanners.

3.1 | Motivation and scope of study

Considering reports in the literature, the goal of this study is to enhance the performance of classical clustering-based segmentation methods and retain all its advantages over the supervised methods. To attain this goal, we introduce the concept of clustering tree to replace the predefined number of clusters required in the classical clustering techniques. A clustering tree provides additional information in the spatial domain. This can render the algorithm robust to images with different levels of contrast, the local optimum traps, tumour heterogeneity, noise and imaging artefact. The proposed method will incorporate human intelligence by coupling the clustering tree to an artificial neural network which incorporates prior knowledge on pathology based on intensity bandwidth, geometric shape, size and boundary.

4 | PROPOSED CLUSTERING TREE-BASED SEGMENTATION

The proposed method is designed to extract brain tumour in MRI data of glioma patients. It is assumed that the MRI data satisfy the following criteria:

1. In the absence of tumour, the constituent anatomic structures can be classified broadly into three classes: white matter, grey matter, and ventricles.

2. In the presence of tumour, the constituent anatomic structures can be classified broadly into four classes: tumour, white matter, grey matter, and ventricles.
3. The tumour (which is the abnormal structure) within the slice is hyperintense relative to the normal structures.
4. The MRI data possess sufficient quality attributes to distinguish between tumours and normal structures.

The first and second criteria underscores the principles underlying the design of clustering algorithms. Clustering algorithms are designed for datasets that contain structures. The third criteria implies that clustering algorithms can be effective in brain MRI only if the abnormal tissue exhibits attributes that distinguish it from the normal structures. The reasoning here is that if the cluster centres are arranged in ascending order, the maximum cluster centre at each clustering resolution should capture the different intensity and geometric attributes of the tumour. The fourth criterion is a statement that the performance of a computer-aided segmentation system should be comparable to the human visual system.

The flow chart in Figure 5a shows that the proposed segmentation method begins with the extraction of low-level features (denoted as LLFE) from the test image, followed by the extraction of high level features (denoted as HLFE) and the classification (denoted as CLFN) of the high level features. Figure 5b shows that the low-level features extracted at the LLFE stage are implemented using the clustering tree CLTT. A schematic for implementing a clustering tree at the clustering resolution $\kappa = n$ is displayed in Figure 4b. Figure 5b also shows that the extraction of high level features at the second stage HLFE is implemented using the single layer feedforward neural network SLNN (see Figure 3b) and that the classification HLFE of the image into tumour and healthy tissues is implemented using the basic model of an artificial neural network BMNN (see Figure 3a). The three successive stages of the implementation process are described in detail using the flow chart in Figure 5c.

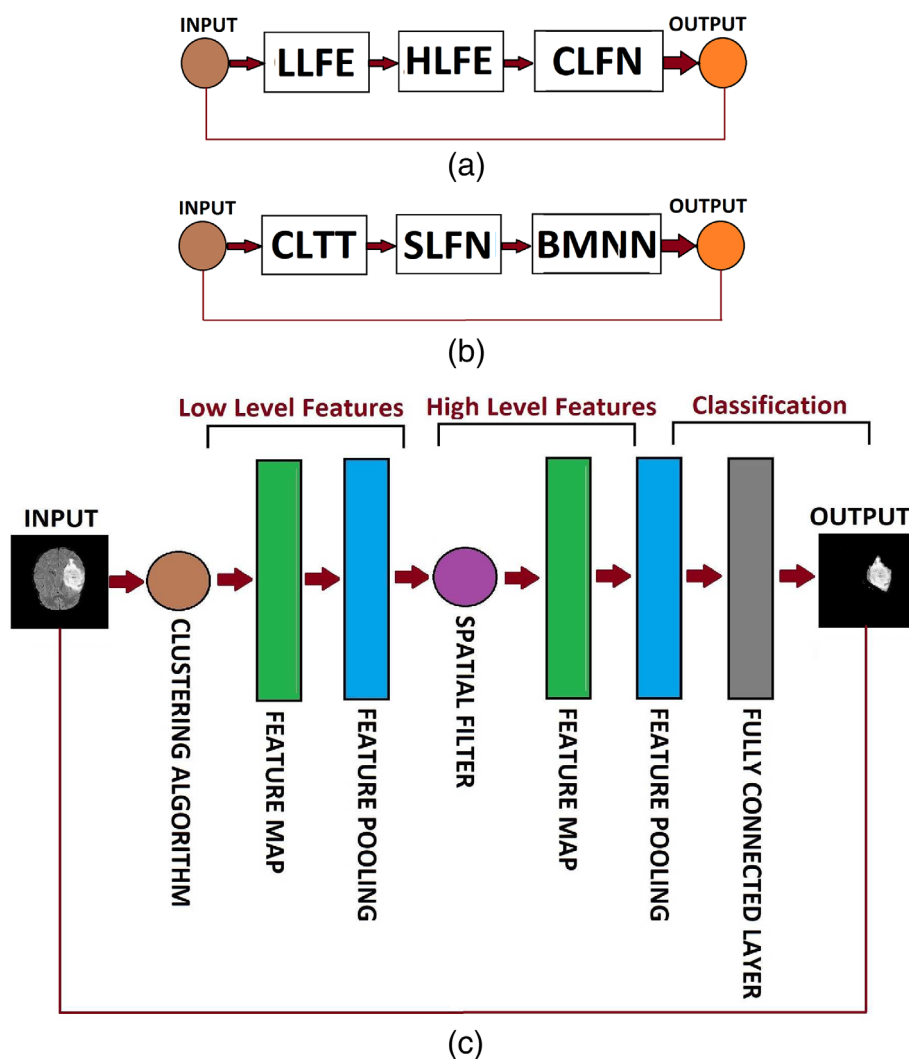


FIGURE 5 Three flow charts that explain the three successive stages in the implementation of the proposed clustering tree-based segmentation method. In each of the flow charts, the verification unit links the output unit with the input unit (a) low-level feature extraction LLFE, followed by high-level feature extraction HLFE and classification CLFN. (b) a clustering tree CLTT is coupled to a single-layer feedforward neural network SLFN, and the output of the SLFN is coupled to a basic model neural network BMNN. (c) Detailed algorithm flow chart

4.1 | Low-level feature extraction (LLFE)

The test image (denoted **TIM**) in Figure 6a is the input to a clustering algorithm that generates a clustering tree **CLTT** shown in Figure 4b, Figure 5b, c, at a clustering resolution of $\kappa = 9$. Each image in Figure 6b–j is a duplicate copy of **TIM** with sets of labels $C_k = \{C_p^k(i, j)\}$. The index

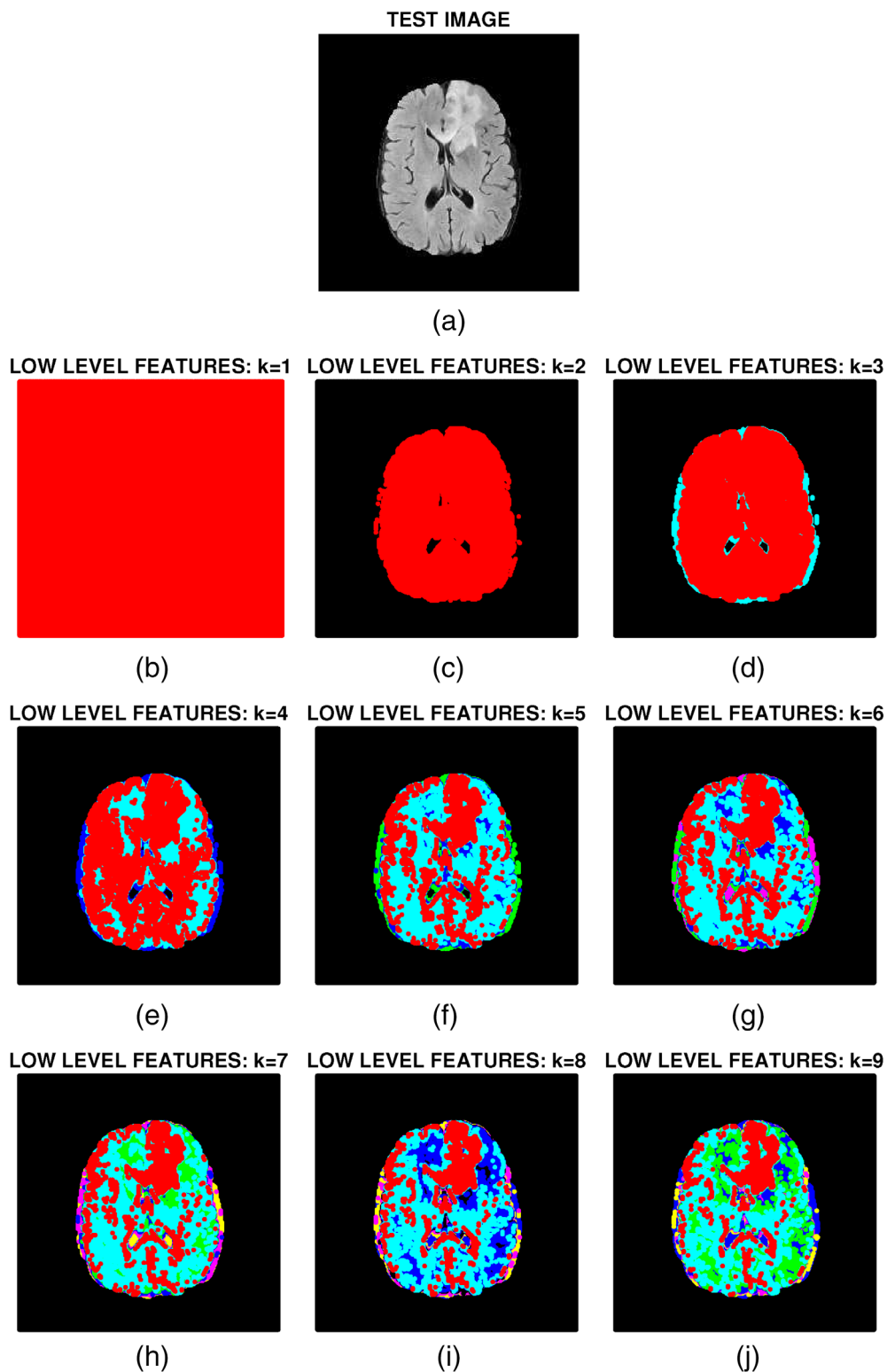


FIGURE 6 Low-level feature images generated from the clustering tree at clustering resolution $\kappa = 9$. The hyperintense pixels (in red) describe the likely locations of the tumour. The red pixels gradually shrink in size as the predefined number of clusters increases

of each duplicated image varies from $p = 1$ to $p = 9$. Feature pooling (denoted as p_k in Figure 4b) is executed on each clustered duplicate copy of TIM. The feature pooling extracts only the cluster $e_{p_k}(i, j)$ in each $C_k = \{C_p^k(i, j)\}$ with the maximum label index:

$$e_{p_k}(i, j) = \arg \max_p \{C_p^k(i, j)\}. \tag{5}$$

The red clusters in Figure 6b-j are those with the maximum label index and are the regions-of-interest. The global threshold (denoted as t_k in Figure 4b), set at k , converts each pooled cluster $e_{p_k}(i, j)$ into a binary image Q_k :

$$Q_k(i, j) = \begin{cases} 1, & e_{p_k}(i, j) = k, \\ 0, & \text{Otherwise} \end{cases} \tag{6}$$

The binary images (denoted as Q_k in Figure 4b) of the pooled low-level features, which are the output of the LLFE stage, are displayed in Figure 7.

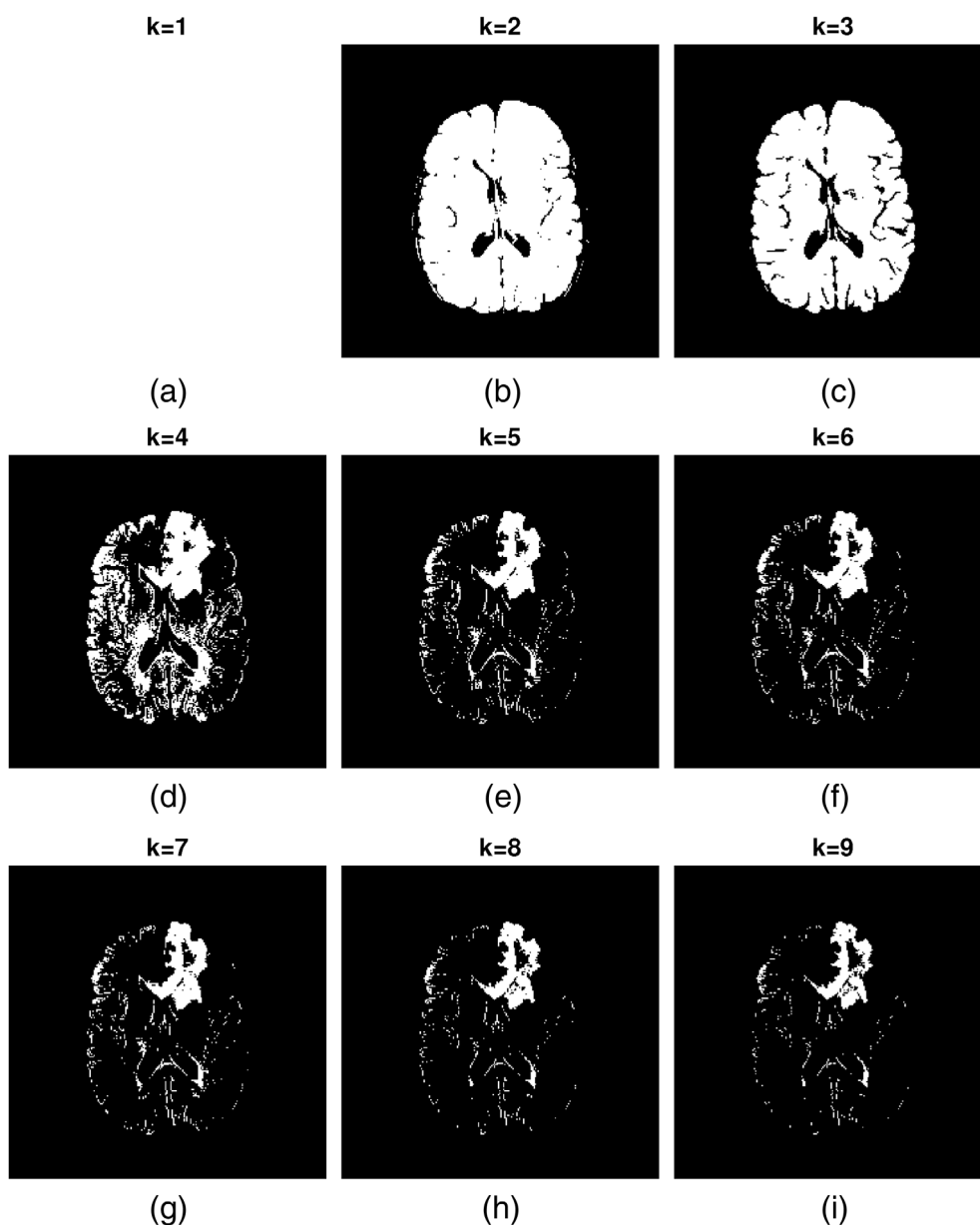


FIGURE 7 The pooled feature image extracted from the low-level feature image. For each clustered duplicate of the original image, the algorithm extracts the cluster with the maximum label index.

4.2 | High-level feature extraction (HLFE)

The output from the LLFE stage is a feature space of dimensions $M \times N \times \kappa$. To ensure that the algorithm of the proposed method is computationally efficient, we follow the recommendation by Übeyli (2008). Accordingly, in a two-step process, we extracted only the relevant features from the feature map. First, beginning from $\kappa = 2$, we compute the sum of absolute difference D_k , an approximation of similarity score between successive feature maps:

$$D_k = \sum_i \sum_j \|Q_k - Q_{k-1}\|_{k=2}^k. \quad (7)$$

The reasoning behind this step is that D_k will have low magnitude where the successive objects have similar characteristics and more likely to contain tumour region. Otherwise D_k will have high magnitude for dissimilar objects of which one is likely to contain non-tumour region. Thus, we select only feature maps $Q_k(i, j)$ where $D_k < 80$ for further processing. Second, we exploit the general characteristics of medical image patterns, which are either circular symmetric or appear as small objects with a variety of geometric patterns (Meyer-Baese & Schmid, 2014).

Images from the LLFE stage comprise objects called connected components and background pixels. Let $\tilde{q}_k^a(i, j)$, where $1 \leq a \leq n_c$ denote the n_c connected components in $Q_k(i, j)$. We regard each $\tilde{q}_k^a(i, j)$ as a neuron. It is the input signal $X_{\tilde{q}_k^a}(i, j)$ to a single layer feedforward neural network (SLFN) (shown in Figure 3b). This input signal is connected to five neurons $\{q_1, q_2, q_3, q_4, q_5\}$ in the hidden layer. The weights $\{w_1, w_2, w_3, w_4, w_5\}$ that connect the input signal with each neuron in the hidden layer correspond to five geometric features. The features are based on size, shape, and boundary to reflect the general characteristics of medical image patterns (Görgel et al., 2015). The five geometric features as defined in Mingqiang et al. (2008) are described below:

1. Area

The area r_1 measures the size of the object:

$$r_1 = \sum_{i=1}^{n_c} \tilde{q}_k^a(i, j), \quad (8)$$

where n_c is the number of bright pixels.

2. Eccentricity

The eccentricity r_2 describes the boundary of an object. It is the ratio of the major axis length z_1 to the minor axis length z_2 of an ellipse with second moment equal to that of the region enclosed by the object.

$$r_2 = \sqrt{1 - \frac{z_1^2}{z_2^2}}. \quad (9)$$

The eccentricity is also a measure of the circularity or ellipticity of an object. An eccentricity that tends to zero (one) indicates that the object is circular (elliptical).

3. Solidity

The solidity r_3 expresses the geometric shape of the object in terms of its degree of convexity:

$$r_3 = \frac{A_c}{H_c} \quad (10)$$

where A_c is the area covered by the contour that describes the object and H_c is the convex hull area of the object. A solidity of one indicates that the connected component has a convex shape.

4. Euler number

The Euler number r_4 measures geometric shape as the relation between the number n_h of contiguous parts and the number n_h' of holes in the object:

$$r_4 = n_h - n_h'. \quad (11)$$

5. Minor axis length

The minor axis r_5 is another boundary descriptor. It is the shortest line through the centre of an ellipse having an equivalent diameter as the object.

According to Equation 2, the output R_l from each of the five neurons (denoted as q_l in Figure 3b) in the hidden layer is

$$R_l = (b_l + W_l X_l) = W_l X_l = r_l \tilde{q}_k^a(i, j), \quad (12)$$

where the bias b_l is assumed to be zero in the neural network and $1 \leq l \leq 5$. The activation function f_l transforms the input to each neuron in the hidden layer according to Equation 2:

$$f_1(R_1(i, j)) = \begin{cases} \tilde{q}_k^a(i, j), & 20 \leq R_1(i, j) \leq 3000, \\ 0, & \text{Otherwise} \end{cases} \quad (13)$$

$$f_2(R_2(i, j)) = \begin{cases} \tilde{q}_k^a(i, j), & R_2(i, j) \geq 0.2, \\ 0, & \text{Otherwise} \end{cases} \quad (14)$$

$$f_3(R_3(i, j)) = \begin{cases} \tilde{q}_k^a(i, j), & R_3(i, j) \geq 0.5, \\ 0, & \text{Otherwise} \end{cases} \quad (15)$$

$$f_4(R_4(i, j)) = \begin{cases} \tilde{q}_k^a(i, j), & -40 \leq R_4(i, j) \leq 3, \\ 0, & \text{Otherwise} \end{cases} \quad (16)$$

$$f_5(R_5(i, j)) = \begin{cases} \tilde{q}_k^a(i, j), & R_5(i, j) \geq 20, \\ 0, & \text{Otherwise} \end{cases} \quad (17)$$

Each activation function f_l is the equivalent of a spatial filter (see Figure 5c), which extracts and propagates only connected components with geometric properties that best describe a brain tumour. The output from each f_l are processed by another activation function $g(i, j)$, such that the output of the SLFN is expressed according to Equation (3):

$$Y(i, j) = \begin{cases} \tilde{q}_k^a(i, j), & \forall f_l(i, j) = \tilde{q}_k^a(i, j), \\ 0, & \text{Otherwise} \end{cases} \quad (18)$$

Equation (18) above implies that the activation function $g(i, j)$ ensures that only connected components that satisfy all the five geometric features are propagated. The neurons from the SLFN are propagated onto a 2D grid having row and column dimensions identical to those of the test image. The pixel locations of the propagated neurons are identical to those of the neurons in their previous 2D grid. Figure 8 shows the output of the SLFN after it processes each object in each binary image corresponding to the index of the duplicated copies of the original image. Theoretically, it can be considered that the SLFN retains the background pixels and converts all connected components that are outside the limit of geometric features defined for brain tumours, into background pixels.

4.3 | Classification

The output from the HLFE stage is the input signal to the basic model of an artificial neural network BMNN shown in Figure 3a. Unlike the HLFE stage, each pixel at the classification stage is regarded as both a neuron and a vector in an $M \times N \times \kappa$ feature map. The value of each input signal to the BMNN is equal to the pixel intensity level: zero or one. The connecting weights between each input signal and the neuron in the BMNN is a constant value of one. The output Y from the BMNN is a 3D image computed according to Equation 1. The output from the BMNN is propagated to a Fully Connected Layer. At the fully connected layer, each neuron is summed along the third dimension:

$$\tilde{Y}(i, j) = \sum_{k=1}^9 Y(i, j, k). \quad (19)$$

The summation along the third dimension is followed by global threshold set at $\tau = 0$ according to Equation (2), resulting in the binary image shown in Figure 9a. The binary image is moved to the output unit of the system.

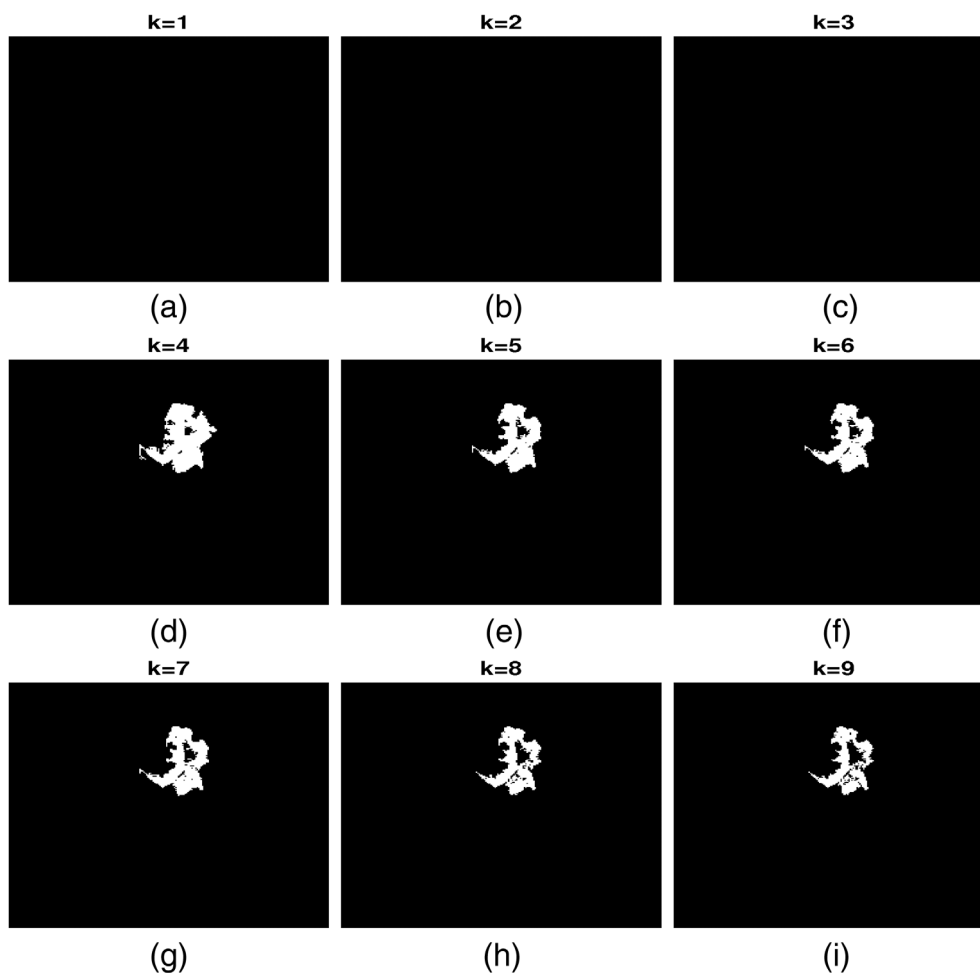


FIGURE 8 High-level feature images extracted from the pooled low-level features. For each pooled low-level feature image, the high-level features were extracted based on the shape, size, and boundary, which generally describe a brain tumour.

4.4 | Output

There are two self-evaluation units within the output of the algorithm. In the first unit, the algorithm evaluates itself by processing the output of the fully connected layer. The flow chart in Figure 9b describes how the self-evaluation is executed. The output from the fully connected layer (denoted as **SEG1**) is regarded as the preliminary output of the algorithm. The function (denoted **IND**) identifies the indices of the bright pixels contained in the tumour region described by **SEG1**. Thereafter, the function **MEN** utilizes the indices of the identified pixels to compute the mean μ_o and standard deviation σ_o of the tumour region in the test image **TIM**. The function **GTR** sets a global threshold at $T_o = \mu_o - \sigma_o$, which is used to produce a binary image (denoted **BIM**) from the test image **TIM**. The function **DCE** computes the Jaccard similarity score $J(\text{TIM}, \text{SEG1})$ between **BIM** and the preliminary segmented image **SEG1**:

$$J(\text{TIM}, \text{SEG1}) = \frac{|I_{\text{BIM}} \cap I_{\text{SEG1}}|}{|I_{\text{BIM}} \cup I_{\text{SEG1}}|}. \quad (20)$$

The function **GTD** generates an image **DCG** that has row and column dimensions identical to those of the original image **TIM**. The image **DCG** is one of two types depending on the threshold set for $J(\text{BIM}, \text{SEG1})$. If $(J(\text{BIM}, \text{SEG1}) > 0.3)$, **DCG** has pixel values of only one, an indication of acceptance of the segmented image. Otherwise, **DCG** will have only zero pixel values thereby rejecting the segmented image. The function **FSG** multiplies the preliminary output **SEG1** with **DCG** to obtain the final output (denoted **SEG**) of the algorithm. The second self-evaluation unit evaluates the output of the algorithm only when the first self-evaluation unit indicates an acceptance of the segmented image. This unit counts the number N_{SEG} of segmented tumours. If $(N_{\text{SEG}} < 2)$ the algorithm finally accept the segmented image, for $(2 < N_{\text{SEG}} < 6)$ only the tumour with maximum size is selected and the algorithm will output only zero pixels if $(N_{\text{SEG}} > 6)$.

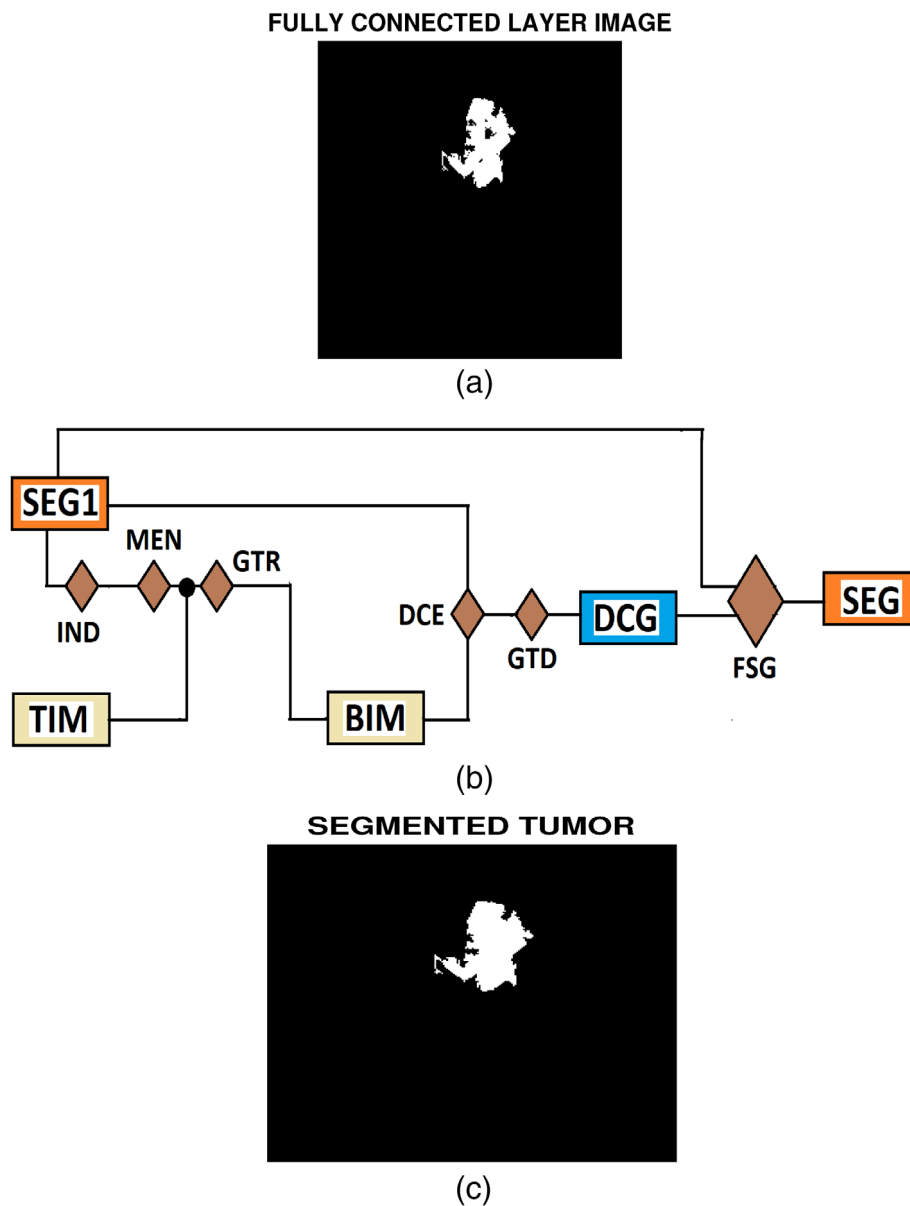


FIGURE 9 (a) At the fully connected layer, the preliminary segmented image is determined by summing pixels along the third dimension of the high-level feature map, followed by global threshold. (b) Flow chart of self-evaluation unit in the output of the algorithm. (c) the segmented tumour.

The self-evaluation unit models the feedback on human behaviour (Panattil et al., 2021). Its design philosophy assumes that the tumour is segmented correctly only if a binary image derived from the test image **TIM** using pixel information from **SEG1** demonstrates a reasonable degree of similarity with **SEG1**. The self-evaluation unit can be considered as a constrained form of feedback because its output does not result in the reinforcement of the algorithm performance. The self-evaluation unit enables the algorithm to detect a poorly segmented image in real time and reduce false positive detection of tumour caused by low image quality or violation of the criteria outlined at the beginning of this section. There is risk of false positive detection when the slices violate the first and second criteria. For example, the inferior and superior slice images can be considered to belong to the two-tissue class rather than the required three-tissue class, particularly when tumour is absent on the slices.

5 | EXPERIMENTS AND RESULTS

The proposed segmentation methods are based on four classical clustering algorithms: k-means, expectation maximization, fuzzy c-means and Otsu. The four proposed methods are referred to as k-means-based, expectation maximization-based, fuzzy c-means-based and Otsu-based

clustering trees. These segmentation methods were implemented in the MATLAB computing environment. The MATLAB implementation code is available on request from the website (<https://www.ntnu.edu/web/colourlab/software>).

5.1 | Sources and description of data

The proposed method was evaluated on 87 MRI volumes from the BRATS2015 dataset and 100 MRI volumes from the BRATS2020 dataset (Menze et al., 2014). The BRATS is an annotated training dataset used for the brain tumour segmentation challenge organized by the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). There are several reasons for which we adopt the BRATS dataset for evaluating the proposed method. First, because the data were acquired from different clinical trial sites, it potentially contains the different characteristics of MRI images that can be observed in a clinical setting. Second, it is the benchmark for evaluating most state-of-the-art brain segmentation algorithms. Third, the dataset contains anonymised clinical data as well as the ground truth, which were annotated by a trained expert. Fourth, the BRATS data had undergone significant preprocessing, for example skull stripping. This makes it amenable to segmentation.

Each patient data has four MRI modalities (T1, T2, T1c and FLAIR), which are aligned to the same anatomical template. This characteristic enables the exploitation of the advantages of multispectral data. However, we utilize only the FLAIR MRI sequence for evaluation because its unique characteristics makes it suitable for highlighting tumour boundary, tumour classification, radiation planning, and treatment response (Soltaninejad et al., 2017). Each MRI volume in both BRATS datasets have dimensions $240 \times 240 \times 155$. We eliminated 29 and 35 most inferior and superior slices, respectively, in each MRI volume because they do not contain any structure or have limited structural information. Thus, only 91 of the 155 slices are considered in each MRI volume. This gives a total of 7917 and 9100 slices from all the MRI volumes from both datasets and we compute and analyse six physical characteristics of tumours across slices in their MRI volumes. They are area, circularity, perimeter, diameter, number of tumours and location. The analysis shows that the interquartile range of the tumour sizes are from few pixels to less than 2500 pixels with median size close to 800 pixels and approximately 100 tumours larger than 2500 pixels are outliers (Figure 10a). The tumours generally exhibit a circular shape because the interquartile range of tumour circularity is between 0 and 2 pixels with median of 0.5 pixels (Figure 10b). The perimeters of most tumours are less than 300 pixels with median value less than 200 pixels (Figure 10c) while the maximum and median tumour diameter are 60 pixels and 37 pixels, respectively (Figure 10d). Most slices contain either a single or double tumours (Figure 10e) with most tumours located in the central region of the human head (Figure 10f).

5.2 | Preprocessing

Although images in the BRATS database had undergone significant preprocessing, there was scope for further preprocessing to maximize the output of the segmentation process. Image quality of each slice extracted from the 3D volume data of a patient was evaluated, based on contrast and noise quality factors, using the algorithms proposed by Osadebey et al. (2017) and Coupé et al. (2010), respectively. If the contrast quality factor was below 0.45, the contrast quality was enhanced by adapting the framework proposed in Sdiri et al. (2016) to the MRI images. The algorithm proposed by Manjón et al. (2015) was applied to reduce noise when the Rician noise level was higher than 5% of the maximum intensity in the image. The images were evaluated for the presence of bias fields, and when necessary, were corrected for intensity inhomogeneity by using the method proposed in Tustison et al. (2010).

5.3 | Experiments

There are three sessions of the performance evaluation. The first session utilize only a subset of the experiment data. Three slices from each of the inferior (slice numbers 30, 40 and 50), central (slice numbers 78, 79 and 80), and superior (slice numbers 100, 110 and 120) sections of the brain were extracted from each of the 87 MRI volume data. This amounted to a total of 783 slice images. The second session utilize all the useful 91 slices in each 87 and 100 MRI volumes from the BRATS2015 (total of 7917) and BRATS2020 datasets (total of 9100), respectively, to evaluate the four proposed clustering tree algorithms based on four criteria. They are overall segmentation performance, true negative detection, tumour size and tumour location. Tumour sizes were categorized as small (less than 700 pixels), medium (between 700 and 1400) and large (greater than 1400 pixels). Tumour locations are inferior (slice numbers 30–60), central (slice numbers 61–90) and superior (slice numbers 91–120).

The third session evaluates the performance of UNet and SegNet deep learning techniques for comparison with the proposed method. Sixty seven and 75 MRI volumes from the 87 and 100 MRI volumes in BRATS2015 and BRATS2020 datasets, respectively, was used for training and

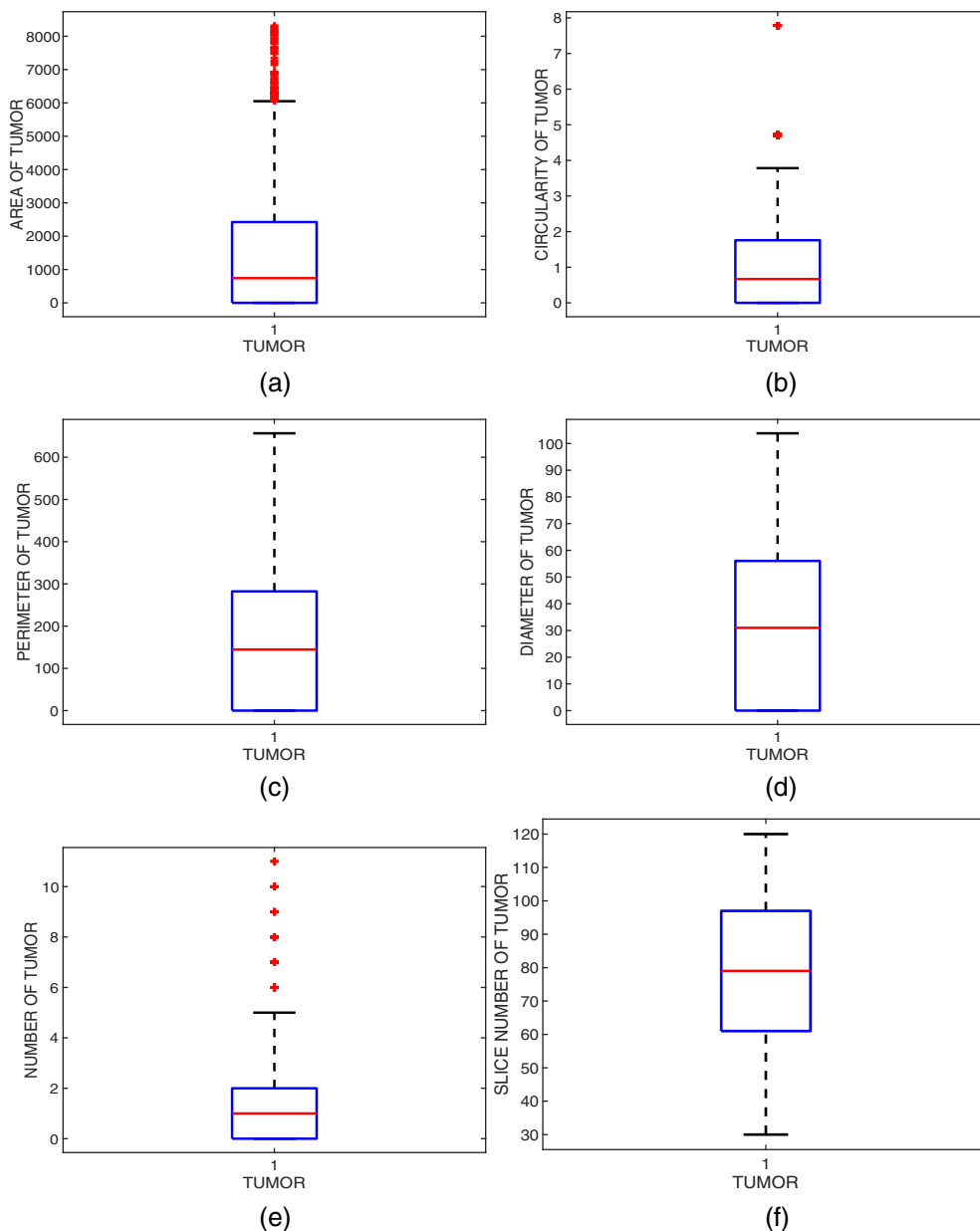


FIGURE 10 Boxplot distribution of physical characteristics of tumours across slices in MRI volumes of glioma patients (a) area, (b) circularity, (c) perimeter, (d) diameter, (e) number of tumours, and (f) location

the remaining corresponding 20 and 25 MRI volumes were used for testing. The input image were preprocessed in same manner as we did in the evaluation of the proposed method except that the data is normalized by subtracting each MRI volume with its mean and dividing it by its standard deviation. The input preprocessed image is of size $240 \times 240 \times 3$, where the FLAIR image is replicated trice to match the three channels of both networks. Both networks had an encoder-decoder depth of 3 where each block contains a 3×3 convolutional kernels with padding set to give same size as the input image. At each training session, the mini-batch size, learning rate and gradient threshold was set to 80, 0.01 and 1, respectively. The network parameters were updated using stochastic gradient descent with momentum optimizer. Data augmentation was performed to increase training data and improve model accuracy by rotating, flipping and shifting slices.

In each experiment, the segmented image I_{seg} was evaluated with reference to its corresponding ground truth I_g based on the Dice similarity coefficient D (Pietka et al., 2010):

$$D = \frac{2 |I_{seg} \cap I_g|}{|I_{seg}| + |I_g|} \quad (21)$$

6 | RESULTS

Evaluation of each proposed clustering tree-based method and its corresponding classical clustering algorithm at clustering resolutions varying from $\kappa = 1$ to $\kappa = 20$ at unit step intervals are presented in Figure 11a–d and Table 1. Comparative performance evaluation of the four proposed clustering tree-based methods on 87 and 100 MRI volumes from BRATS2015 and BRATS2020 datasets at clustering resolutions varying from $\kappa = 1$ to $\kappa = 20$ are presented in Tables 2 and 3, respectively. We combined the performance evaluation results on 187 MRI volumes from the two BRATS datasets and present their statistics in Figure 12. The comparative performance evaluation results between the four proposed clustering-tree based methods and the two deep neural network algorithms (UNet and SegNet) on 20 and 25 MRI volumes from the BRATS2015 and BRATS2020 datasets are displayed Tables 4 and 5, respectively.

There is a common performance trend among the four proposed algorithms. As the clustering resolution increases, there is a corresponding increase in the performance of each proposed clustering algorithm. However, this common performance trend attained saturation at different clustering resolutions. Figure 11a shows that the performance of the expectation maximization-based clustering tree method attained a peak Dice score of 0.72 at $\kappa = 6$. Figure 11b shows that the performance of the fuzzy c-means-based clustering tree method attained a peak Dice score of 0.84 at $\kappa = 6$. The performance of the k-means-based clustering tree method was saturated at $\kappa = 6$, when the Dice score was 0.88 (see Figure 11c). Meanwhile, the performance of the Otsu-based clustering tree method was saturated at $\kappa = 7$, when the Dice score was 0.87 (see Figure 11d).

Each of the four proposed methods and their corresponding classical techniques demonstrate an equal level of performance at lower clustering resolutions and at predefined number of clusters, respectively. The proposed expectation maximization-based clustering tree method and its corresponding classical technique demonstrate an equal level of performance for the clustering resolution ($1 \leq \kappa \leq 3$) and predefined number of clusters ($1 \leq k \leq 3$) (see Figure 11a). A similar trend was exhibited by the proposed fuzzy c-means-based clustering tree method and its corresponding classical technique for ($1 \leq \kappa \leq 4$) and ($1 \leq k \leq 4$) (see Figure 11b). The proposed k-means-based clustering tree method and its corresponding classical technique also exhibited a similar performance trend for ($1 \leq \kappa \leq 3$) and ($1 \leq k \leq 4$) (see Figure 11c). The proposed Otsu-based clustering tree method and its corresponding classical technique also exhibited a similar performance trend for ($1 \leq \kappa \leq 4$) and ($1 \leq k \leq 4$) (see Figure 11d). Beyond the region of equal performance trends, the performance of the classical technique reduces gradually, whereas the performance of the proposed method increases up to a saturation point. The additional strength exhibited by the four proposed methods over their corresponding classical techniques was anticipated owing to the additional and valuable information provided by the clustering tree.

The comparative performance evaluation results from the first experiment shown in Figure 11e shows that the Otsu-based and k-means-based clustering tree methods, with a mean Dice score of 0.87 and 0.88, respectively, demonstrates better performance over the other two proposed algorithms. The fuzzy c-means-based clustering tree method comes next with a mean Dice score of 0.84. The expectation maximization-based clustering tree method trailed behind other proposed techniques with a mean Dice score of 0.72. The fuzzy, k-means and Otsu techniques recorded Dice scores of 0.84, 0.88 and 0.85, respectively. These Dice scores quantifies the robustness of the four proposed algorithms to images with different quality variations. The standard deviation Dice score of 0.06 recorded by the k-means-based clustering tree method reveals it to be the most robust of the four proposed algorithms. As anticipated, the proposed expectation maximization-based clustering tree method trailed behind the other two proposed methods. This may be partly owing to the Gaussian assumption in the formulation of the EM algorithm whereas the pixels in MRI images follow the Rician distribution.

In the second experiment that utilize all the experiment data, all the proposed methods achieved comparable median Dice scores on the overall performance criteria (Figure 12a). However, a closer look at the inter quartile range of the boxplot shows that the distribution of Dice score for k-means and Otsu is concentrated within the high Dice score spectrum whereas the expectation maximization and fuzzy have Dice score distribution across all range of values of the Dice score (Tables 2 and 3). The k-means recorded most concentration of high Dice scores with low Dice scores as outliers. The range of true negative detection for the proposed methods are from 0.60 to 0.9 with k-means demonstrating the best performance followed by Otsu method (Figure 12b). All the proposed methods demonstrate remarkable segmentation performance for medium and large sizes of tumours. However, for small tumours, the k-means technique achieved only an average performance whereas the expectation maximization and fuzzy perform poorly (Figure 12c). The k-means and Otsu demonstrate robust performance for tumours across different regions of the brain. Expectation maximization and fuzzy clustering perform dismally at inferior region. Moreover, fuzzy perform poorly at central region of the brain. On single, double and multiple tumours, all the proposed methods, except fuzzy, demonstrate same level of performance (Figure 12d). Therefore, we can conclude that the fuzzy technique has poor true positive detection but very good true negative detection.

In the third experiment session, the performance of the proposed methods based on k-means and Otsu techniques can be said to be comparable to UNet and SegNet deep learning techniques. The UNet algorithm had the best performance of 0.95 median Dice score based on the BRATS2015 dataset. The UNet was closely followed by the SegNet with median Dice score of 0.93, Otsu recorded mean Dice score of 0.91 and k-means has mean Dice score of 0.90. The UNet and SegNet were also the best performing algorithms on the BRATS2020 dataset and were closely followed by Otsu and k-means algorithms.

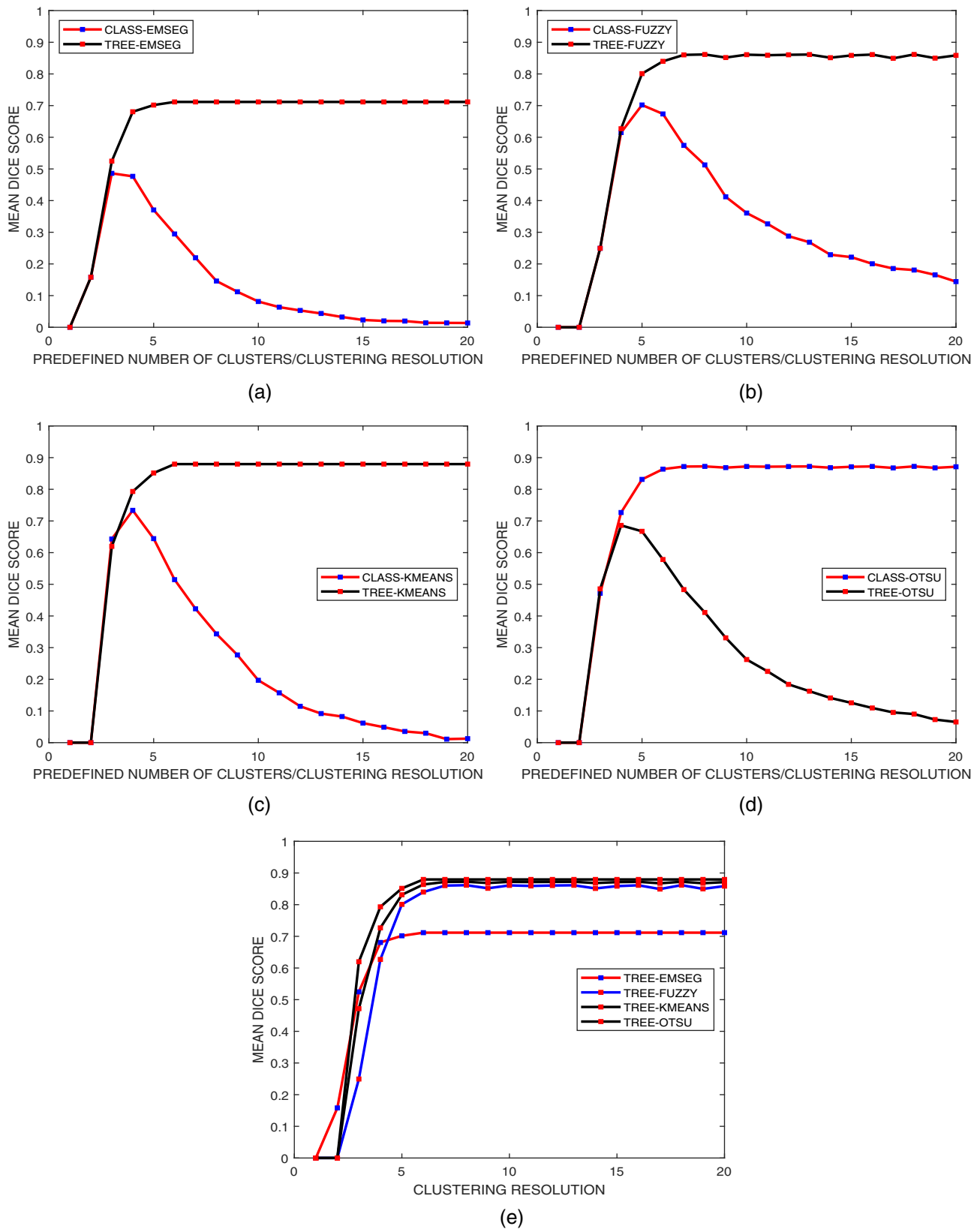


FIGURE 11 (a) Average dice score-based comparative performance evaluation of the four classical clustering tree-based segmentation methods for clustering resolutions varying from $\kappa = 1$ to $\kappa = 20$, and their corresponding classical clustering-based methods for predefined numbers of clusters varying from $k = 1$ to $k = 20$ (a) expectation maximization, (b) fuzzy c-means, (c) k-means, (d) Otsu and (e) average dice score-based comparative performance evaluation of the four clustering tree-based segmentation methods for clustering resolutions varying from $\kappa = 1$ to $\kappa = 20$

TABLE 1 Comparative performance evaluation of the four classical clustering and their corresponding clustering tree-based algorithms on a subset of the experiment data

Algorithms	Classical clustering		Clustering tree	
	Mean dice score	Standard deviation dice score	Mean dice score	Standard deviation dice score
Expectation maximization	0.59	0.30	0.72	0.24
Fuzzy	0.73	0.28	0.84	0.15
k-means	0.79	0.24	0.88	0.06
Otsu	0.68	0.21	0.85	0.12

Note: The algorithms were evaluated on three slices from each of the inferior (slice numbers 30, 40 and 50), middle (slice numbers 78, 79 and 80), and superior (slice numbers 100, 110 and 120) sections of 87 brain MRI volume data. This amounted to a total of 783 slice images.

TABLE 2 Comparative performance evaluation of the four clustering tree-based algorithms on 87 MRI volumes from the BraTs 2015 dataset

Algorithms	Dice score			
	Average	Median	Lower quartile	Upper quartile
Expectation maximization	0.57	0.86	0.00	0.96
Fuzzy	0.56	0.88	0.00	1.00
k-means	0.80	0.90	0.72	1.00
Otsu	0.75	0.91	0.52	1.00

TABLE 3 Comparative performance evaluation of the four clustering tree-based algorithms on 100 MRI volumes from the BraTs 2020 dataset

Algorithms	Dice score			
	Average	Median	Lower quartile	Upper quartile
Expectation maximization	0.62	0.83	0.41	1.00
Fuzzy	0.70	0.90	0.46	1.00
k-means	0.81	0.93	0.74	1.00
Otsu	0.83	0.89	0.61	1.00

7 | DISCUSSION

Although the classical clustering algorithms have highly attractive features, the attainment of a satisfactory level of performance is limited by the presence of noise, tumour heterogeneity, and variation in quality attributes across images in clinical settings. There is presently no widely acceptable objective approach to determining the number of clusters required to achieve optimal segmentation because most classical clustering techniques exhibit cluster instability and are vulnerable to the traps of local minima. This study proposes a new approach to enhancing the performance of the classical clustering algorithms. We describe below the characteristics of the proposed method that makes it effective for data scientists and clinical practitioners such as radiologists, neurologists, and pathologists. Thereafter, potential applications of the proposed method, its limitations, and future research direction are highlighted.

7.1 | Algorithm design and operational features

1. Enhanced performance over classical clustering methods

The performance evaluation results displayed in Figure 11 and Table 1 reveal that the proposed method enhances the performance of the classical clustering-based segmentation methods. The mean Dice score recorded by the proposed k-means-based clustering tree, expectation maximization-based clustering tree, fuzzy c-means-based, and Otsu-based clustering tree techniques are 0.88, 0.72, 0.84 and 0.85, respectively, whereas those recorded by the corresponding classical clustering techniques are 0.79, 0.59, 0.73, and 0.68, respectively. The standard deviation

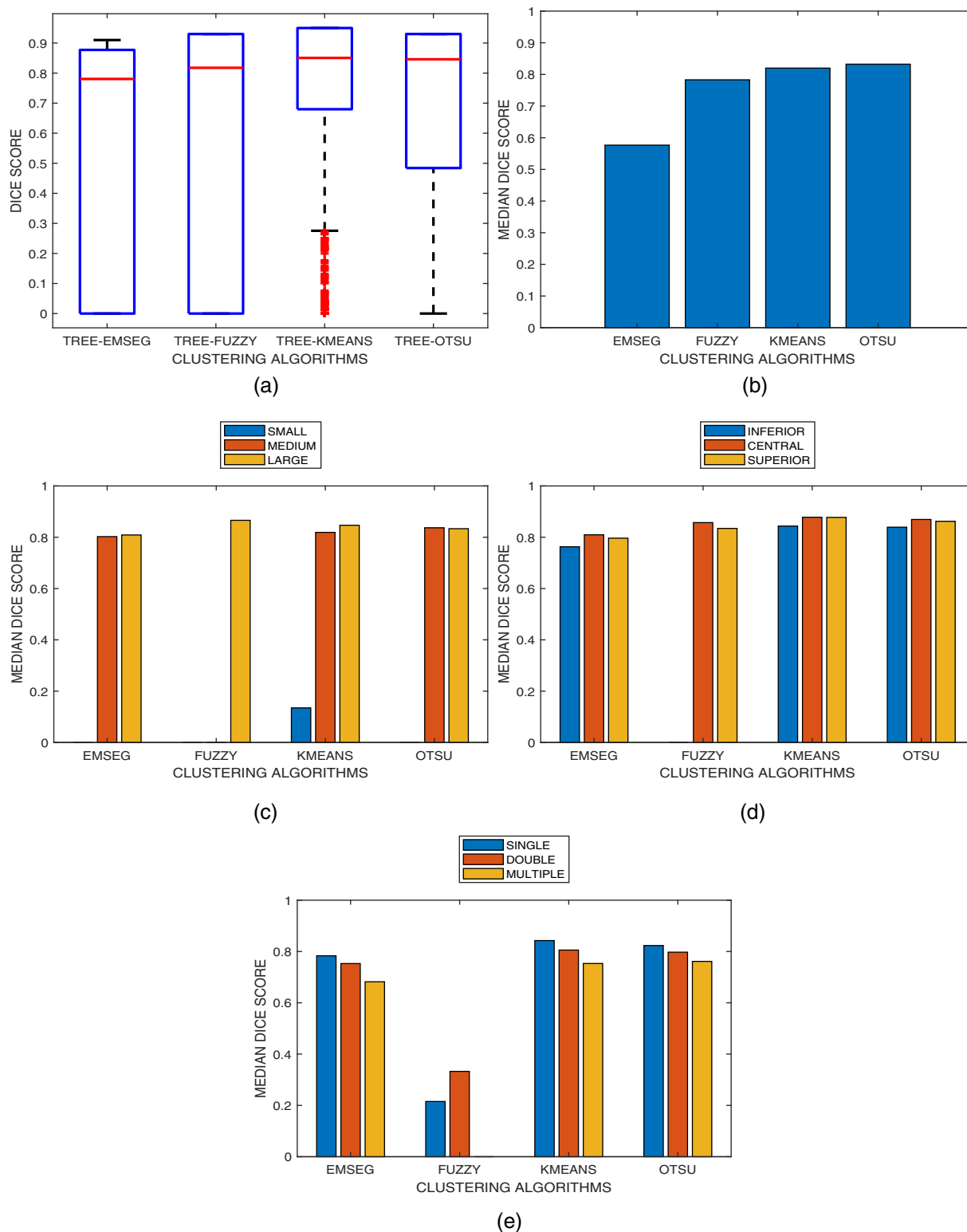


FIGURE 12 (a) Comparative performance evaluation of the four classical clustering tree-based segmentation methods at clustering resolutions varying from $\kappa = 1$ to $\kappa = 20$. The algorithms were evaluated on all useful slices across 87 and 100 glioma patients MRI volumes in the BraTs 2015 and 2020 datasets, respectively. Comparative performance evaluation of the four classical clustering tree-based segmentation methods based on five tumour attributes (b) none-tumorous slices, (c) tumour sizes, (d) tumour locations, and (e) number of tumours, for clustering resolutions varying from $\kappa = 1$ to $\kappa = 20$

TABLE 4 Comparative performance evaluation of the four clustering tree-based algorithms and two CNN-based deep learning algorithms on 20 MRI volumes from the BraTs 2015 dataset

Algorithms	Dice score			
	Average	Median	Lower quartile	Upper quartile
Expectation maximization	0.71	0.86	0.51	1.00
Fuzzy	0.81	0.88	0.63	1.00
k-means	0.84	0.90	0.72	1.00
Otsu	0.86	0.91	0.71	1.00
UNet-CNN	0.88	0.95	0.77	1.00
SegNet-CNN	0.87	0.93	0.75	1.00

TABLE 5 Comparative performance evaluation of the four clustering tree-based algorithms and two CNN-based deep learning algorithms on 25 MRI volumes from the BraTs 2020 dataset

Algorithms	Dice score			
	Average	Median	Lower quartile	Upper quartile
Expectation maximization	0.73	0.82	0.60	1.00
Fuzzy	0.79	0.83	0.66	1.00
k-means	0.78	0.85	0.76	1.00
Otsu	0.80	0.90	0.74	1.00
U-net CNN	0.83	0.92	0.79	1.00
SegNet-CNN	0.81	0.93	0.73	1.00

of the Dice scores recorded from the clustering tree techniques are 0.06, 0.24, 0.15 and 0.12, respectively, whereas those recorded by the corresponding classical clustering methods are 0.24, 0.30, 0.28 and 0.21, respectively.

2. Incorporation of anatomical knowledge

The incorporation of anatomical knowledge into segmentation algorithms is highly desired for automation in expert systems (Gordillo et al., 2013). The proposed method satisfies this design criterion by incorporating five separate spatial filters. The parameter for each filter is tuned to extract five geometrical parameters: area, eccentricity, solidity, Euler number, and minor axis length. This approach discriminates between healthy and abnormal anatomical structures effectively. Furthermore, the clustering tree concept in the proposed method exploits the general knowledge on pathology based on intensity attributes, to maximize the segmentation outcome.

3. Guide to determine optimal number of clusters

The accuracy of the clustering-based segmentation method is strongly dependent on the proper incorporation of prior knowledge through the selection of optimal segmentation parameters. Because there is no widely acceptable objective method to compute the parameters, heuristics is presently the popular approach. This hinders accurate segmentation (Kwon et al., 2017). The results from this study can guide a data scientist to determine the optimal number of clusters that can maximize segmentation outcomes. The plots in Figure 11a–d show that the predefined number of clusters that can maximize segmentation outcomes for the classical clustering algorithms are $K=3$, $K=5$, $K=4$, and $K=4$ for the expectation maximization, fuzzy c-means, k-means and Otsu-based methods, respectively. The plots in Figure 11e reveal that the clustering resolutions $\kappa=4$, $\kappa=5$, $\kappa=6$, and $\kappa=6$ optimizes segmentation for the expectation maximization-based, fuzzy c-means-based, k-means-based and Otsu-based clustering tree techniques, respectively. Similar to the contribution by Limwattanapibool and Arch-int (2017), the proposed method can guide data scientists and clinical practitioners to accurately predict the predefined number of cluster that will maximize segmentation outcomes in clustering-based expert systems.

4. Attractive features from clustering

Similar to scale space analysis, the concept of clustering trees provides more detailed information within and across anatomical structures in brain MRI images. The multiple images associated with the clustering tree potentially provide spatial information and weaken the effect of noise during the process for extracting tumour.

5. Softening of hard clustering algorithm

The hard clustering procedure in the classical k-means clustering makes it unsuitable for data with a significant level of vagueness. The concept of clustering tree in the proposed method when applied to the classical k-means clustering softens its hard clustering procedure and thereby enhances its performance.

6. Fully unsupervised and fully automated

Most contributions, particularly the semi-supervised learning techniques, are aimed at enhancing the performance of the classical clustering-based methods of segmentation by using minimal labelled training data. This imparts it the advantages of both the supervised and unsupervised techniques. This approach causes the enhanced algorithm to lose its simplicity and feature of minimal user interaction, rendering it unattractive to clinicians. The proposed method is both fully unsupervised and fully automated. The clustering tree was coupled with an artificial neural network that does not require backpropagation for learning with training data. The concept of clustering tree in the proposed method provides additional information, which potentially serves a similar purpose as the use of minimal labelled training data, and ensures that the proposed method retains the simplicity and feature of minimal user interaction of the classical clustering algorithms. At the high-level feature extraction stage, which was implemented using a **SLFN**, the connected components from the output of the clustering tree were modelled as neurons. The weights connecting the single neuron in the input layer and the five neurons in the hidden layer of the **SLFN** was modelled by the geometric features of the connected components. We consider this as another approach to ensuring that the algorithm retains its feature of being fully unsupervised. Another novel approach to ensuring a fully unsupervised operation is the classification stage. Herein, the pixels were regarded as input signal to a **BMNN** and also as vectors in a 3D feature space. The weights connecting each input signal and the individual neuron in the **BMNN** is set to a constant value of one.

7. Features of an expert system

A clustering tree-based method of segmentation display similar design features as tree-based methods of classification because it computes an output based on several inputs. It has been demonstrated that tree-based classification methods demonstrate better performance over other classification methods in the diagnosis and prediction of diseases and content-based image retrieval systems (Dimitrovski et al., 2016; Jha et al., 2019).

The clustering tree (in the low-level feature extraction stage of the algorithm) is conceptually similar to the scale-space theory. There is a close link between the scale space theory and biological vision. The clustering tree was coupled to an artificial neural network that models how humans learn to recognize objects. The self-evaluation unit within the output of the algorithm, which provides feedback that enables the algorithm to evaluate itself, is equivalent to human reasoning. Thus, the proposed method can be considered to exhibit the features of an expert system.

8. Flexibility in design

In the proposed method, we utilize the basic model of a neural network that has only one hidden layer, with a fixed set of parameters for the activation functions. However, depending on the application, the parameters of the activation functions and the number of hidden layers can be varied. Increasing the number of hidden layers is equivalent to having multiple spatial filters in cascade. Thus, the design of the proposed method provides the flexibility to optimize its performance for a variety of segmentation tasks.

9. Interpretability and transparency

The self-evaluation unit coupled with the clustering tree in the proposed method is an effective tool for data scientists and clinicians to visualize and explore the different clustering resolutions. Thereby, it facilitates informed decisions such as the need for image quality enhancement and tuning of the algorithm parameters. This design feature also renders the segmentation process transparent and convenient-to-use and its results conveniently interpretable.

10. Resource for implementation of attention mechanism

Features extracted from images at the optimal number of cluster can be used to build a training set consisting of features relevant for classification or segmentation task. The training set can serve as input to a CNN. This approach will eliminate the use of the entire grayscale image thereby reducing training time as it allows the model to learn from only the relevant features. Therefore, the proposed method is potentially a useful resource for the incorporation of attention mechanism into a CNN architecture.

11. The proposed method and extreme learning machine

The proposed method is related to extreme learning machine. The single layer feedforward neural network at the high level feature extraction layer is basic component of an extreme learning machine. The main difference between the two algorithms is that an extreme learning machine can be used to train single layer feedforward neural network.

12. Comparative performance evaluation with deep neural networks

There is no doubt that the comparative performance evaluation results demonstrate the robustness of deep neural network in segmentation tasks. However, it is important to note that our proposed method does not require training images while deep neural networks rely on example images for learning to achieve acceptable segmentation accuracy. The comparative performance evaluation results also indicate that k-means and Otsu clustering-tree techniques can compete with the deep neural network-based algorithms.

7.2 | Other potential applications

A key step in the proposed method for segmenting tumour in FLAIR MRI images is to extract clusters with the maximum label index in each clustered duplicate of the original image. This step in the proposed method can be modified to extract other regions of interest. For example, in a FLAIR sequence, if the feature pooling at the low-level feature extraction stage focuses on clusters with the minimum label index, the hypointense grey matter and the ventricles become the segmented structure. The flexibility of the proposed method can be exploited to segment tumour in multispectral MRI data. Let us consider a case where the proposed method is applied separately to multispectral MRI data that consist of the T1 and T2 sequences. The low-level feature extraction stage of the proposed method will be configured to extract the hypointense grey matter and ventricles in the T1 sequence by extracting clusters with the maximum label index. For the T2 sequence, the low-level feature extraction stage will focus on the hyperintense structures consisting of tumour, grey matter, and the ventricles. Thereafter, the information extracted from both the images can be combined to determine the tumour region.

The proposed method is potentially effective for evaluating image quality. The multiple images generated by the clustering tree can reveal noisy pixels and provide a path for the quantitative evaluation of image contrast. Lung and heart segmentation in X-ray and computed tomography (CT) images and the segmentation of images acquired from digital cameras are other potential applications of the proposed method.

7.3 | Limitations of proposed method and future work

Although the proposed clustering tree-based technique demonstrate significant segmentation accuracy over its corresponding classical techniques, there is still room for improvement. In real clinical studies attaining acceptable segmentation accuracy will be challenging due to wide variations in image quality attributes. For example, in multi-center clinical studies the use of geometric features such as eccentricity and solidity may no longer be an efficient method to discriminate between tumour and non-tumorous objects. In its present form, the proposed method is robust to relatively large single or double lesions. The algorithms may require modification for efficient segmentation of MRI slices that contains multiple and relatively small tumours. The proposed method is robust to noise because of the different levels of details provided by the clustering tree, however, there is a limit to its performance. Quality attributes of the image plays an important role in the performance of the proposed method. We strongly recommend the preprocessing of images to optimize the performance of the proposed method.

Future work will focus on how to address the challenges in the application of the proposed method to brain analysis in multi-center clinical trials. These challenges can be addressed through the use of an extreme learning machine to train the SLNN, located at the HLFE layer, on geometric attributes of tumour and non-tumorous objects, thereby improving the segmentation accuracy of the proposed method. Another approach to address real-world challenges is to integrate the proposed method into a CNN whereby the proposed method set at the optimal number of clusters will generate image features needed for attention mechanism. Future work will also investigate how and why the fuzzy technique perform poorly on tumorous slices and on improving the robustness of the proposed method to segmentation of multiple tumours and low contrast quality images. Other potential applications of the proposed method will be explored and possibly implemented. The utility of the algorithm will also be extended to include the segmentation of relatively large and small abnormal structures such as lymph nodes in CT images of cancer patients and lesions in MRI images of patients with multiple sclerosis.

8 | CONCLUSIONS

Benefits of classical clustering segmentation techniques include simplicity, unsupervised procedures, computational efficiency, cost-effectiveness and minimal user interface. A significant requirement for its operation is that the user must select the number of clusters that the algorithm will generate. It is challenging and inefficient for a user to manually determine the number of clusters that will maximize the output of a clustering-based segmentation algorithm where a large volume of MRI images are processed. This study proposes to enhance the performance of classical clustering-based segmentation techniques through a shift from manual selection of the predefined number of clusters to clustering trees. The proposed clustering tree-based approach to segmentation was developed for four classical clustering algorithms: k-means, expectation maximization, fuzzy c-means and Otsu. The clustering tree is coupled to artificial neural networks that do not require training. The clustering tree provides

different levels of details and incorporates spatial information that is absent in the classical clustering algorithms. The artificial neural networks mimic human intelligence. Comparative performance evaluation results shows that the proposed clustering tree-based methods outperform the classical clustering-based methods. We consider that the proposed method will be attractive to data scientists and clinicians because of its fully unsupervised and fully automated operational characteristics.

ACKNOWLEDGEMENT

This work was supported by the European Research Consortium on Informatics and Mathematics (ERCIM).

CONFLICT OF INTEREST

We have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in BRATS 2015 and BRATS 2020 at <https://www.med.upenn.edu/cbica/>. These data were derived from the following resources available in the public domain: - Center for Biomedical Image Computing & Analytics, <https://www.med.upenn.edu/cbica/>.

ORCID

Michael Osadebey  <https://orcid.org/0000-0002-4681-2958>

Nizar Bouguila  <https://orcid.org/0000-0001-7224-7940>

REFERENCES

- Aggarwal, S., & Singh, P. (2019). Cuckoo and krill herd-based k-means++ hybrid algorithms for clustering. *Expert Systems*, 36(4), e12353.
- Alibolandi, M., Charbgo, F., Taghdisi, S. M., Abnous, K., & Ramezani, M. (2018). Chapter 4: Active targeted nanoscale delivery systems for brain tumor therapeutics. In P. Kesharwani & U. Gupta (Eds.), *Nanotechnology-based targeted drug delivery systems for brain tumors* (pp. 75–110). Academic Press. <https://doi.org/10.1016/B978-0-12-812218-1.00004-X>
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495.
- Bahra, G., & Wiese, L. (2019). Parameterizing neural networks for disease classification. *Expert Systems*, 37(1), e12465.
- Bandyopadhyay, S., Saha, S., Maulik, U., & Deb, K. (2008). A simulated annealing-based multiobjective optimization algorithm: Aмоса. *IEEE Transactions on Evolutionary Computation*, 12(3), 269–283.
- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2–3), 191–203.
- Bittmann, R. M., & Gelbard, R. M. (2007). Decision-making method using a visual approach for cluster analysis problems; indicative classification algorithms and grouping scope. *Expert Systems*, 24(3), 171–187.
- Bouguila, N., Ziou, D., & Vaillancourt, J. (2004). Unsupervised learning of a finite mixture model based on the dirichlet distribution and its application. *IEEE Transactions on Image Processing*, 13(11), 1533–1543.
- Coupé, P., Manjón, J. V., Gedamu, E., Arnold, D., Robles, M., & Collins, D. L. (2010). Robust rician noise estimation for mr images. *Medical Image Analysis*, 14(4), 483–493.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.
- Despotović, I., Goossens, B., & Philips, W. (2015). Mri segmentation of the human brain: Challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, 2015, 1–23.
- Dimitrovski, I., Kocev, D., Loskovska, S., & Džeroski, S. (2016). Improving bag-of-visual-words image retrieval with predictive clustering trees. *Information Sciences*, 329, 851–865.
- Ding, S., Zhao, H., Zhang, Y., Xu, X., & Nie, R. (2015). Extreme learning machine: Algorithm, theory and applications. *Artificial Intelligence Review*, 44(1), 103–115.
- Edjlali, M., Ploton, L., Maurage, C.-A., Delmaire, C., Pruvo, J.-P., Reyns, N., & Leclerc, X. (2019). Intraoperative mri and flair analysis: Implications for low-grade glioma surgery. *Journal of Neuroradiology*, 48(1), 61–64.
- Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Foufo, S., & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing*, 2(3), 267–279.
- Gandomi, A. H., & Yang, X.-S. (2014). Chaotic bat algorithm. *Journal of Computational Science*, 5(2), 224–232.
- Ganesh, M., Naresh, M., & Arvind, C. (2017). Mri brain image segmentation using enhanced adaptive fuzzy k-means algorithm. *Intelligent Automation & Soft Computing*, 23(2), 325–330.
- Gordillo, N., Montseny, E., & Sobrevilla, P. (2013). State of the art survey on mri brain tumor segmentation. *Magnetic Resonance Imaging*, 31(8), 1426–1438.
- Görgel, P., Sertbas, A., & Uçan, O. N. (2015). Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines. *Expert Systems*, 32(1), 155–164.
- Huang, D.-Y., Lin, T.-W., & Hu, W.-C. (2011). Automatic multilevel thresholding based on two-stage otsu's method with cluster determination by valley estimation. *International Journal of Innovative Computing, Information and Control*, 7(10), 5631–5644.
- Işın, A., Direkoğlu, C., & Şah, M. (2016). Review of mri-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, 102, 317–324.

- Jafri, M. Z. M., Abdulbaqi, H. S., Mutter, K. N., Mustapha, I. S., & Omar, A. F. (2017). Measuring the volume of brain tumour and determining its location in t2-weighted mri images using hidden markov random field: Expectation maximization algorithm. In *Digital optical technologies 2017* (Vol. 10335, p. 103351V).
- Jeub, L. G., Sporns, O., & Fortunato, S. (2018). Multiresolution consensus clustering in networks. *Scientific Reports*, 8(1), 1–16.
- Jha, S. K., Pan, Z., Elahi, E., & Patel, N. (2019). A comprehensive search for expert classification methods in disease diagnosis and prediction. *Expert Systems*, 36(1), e12343.
- Kaur, G., & Rani, J. (2016). Mri brain tumor segmentation methods: A review. *Infinite Study*.
- Kaur, N., & Sharma, M. (2017). Brain tumor detection using self-adaptive k-means clustering. In *2017 international conference on energy, communication, data analytics and soft computing (ICECDS)* (pp. 1861–1865).
- Korez, R., Ibragimov, B., Likar, B., Pernuš, F., & Vrtovec, T. (2015). A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation. *IEEE Transactions on Medical Imaging*, 34(8), 1649–1662.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Kwon, B. C., Eysenbach, B., Verma, J., Ng, K., De Filippi, C., Stewart, W. F., & Perer, A. (2017). Clustervision: Visual supervision of unsupervised clustering. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 142–151.
- Li, Z., Liu, G., Zhang, D., & Xu, Y. (2016). Robust single-object image segmentation based on salient transition region. *Pattern Recognition*, 52, 317–331.
- Limwattanapibool, O., & Arch-int, S. (2017). Determination of the appropriate parameters for k-means clustering using selection of region clusters based on density dbscan (srcd-dbscan). *Expert Systems*, 34(3), e12204.
- Lindeberg, T. (2007). *Scale-space* (pp. 2495–2504). Wiley Encyclopedia of Computer Science and Engineering.
- Lindeberg, T. (2015). Image matching using generalized scale-space interest points. *Journal of Mathematical Imaging and Vision*, 52(1), 3–36.
- Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Soffietti, R., von Deimling, A., & Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro-Oncology*, 23(8), 1231–1251.
- Lu, S., Wang, S.-H., & Zhang, Y.-D. (2020). Detection of abnormal brain in mri via improved alexnet and elm optimized by chaotic bat algorithm. *Neural Computing and Applications*, 33, 1–13.
- MacQueen, J., et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).
- Manjón, J. V., Coupé, P., & Buares, A. (2015). Mri noise estimation and denoising using non-local pca. *Medical Image Analysis*, 22(1), 35–47.
- Mazurek, J., & Mazurek, J. (2013). Examination of european union economic cohesion: A cluster analysis approach. *Business and Economic Horizons*, 9(4), 8–17.
- Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., ... Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2024.
- Merzban, M. H., & Elbayoumi, M. (2019). Efficient solution of otsu multilevel image thresholding: A comparative study. *Expert Systems with Applications*, 116, 299–309.
- Meyer-Baese, A., & Schmid, V. J. (2014). *Pattern recognition and signal analysis in medical imaging*. Elsevier.
- Mingqiang, Y., Kidiyo, K., & Joseph, R. (2008). A survey of shape feature extraction techniques. In *Pattern recognition techniques, technology and applications*. IntechOpen.
- Nalepa, J., Marcinkiewicz, M., & Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13, 83.
- Nazar, U., Khan, M. A., Lali, I. U., Lin, H., Ali, H., Ashraf, I., & Tariq, J. (2020). Review of automated computerized methods for brain tumor segmentation and classification. *Current Medical Imaging*, 16(7), 823–834.
- Osadebey, M., Pedersen, M., Arnold, D., & Wendel-Mitoraj, K. (2017). No-reference quality measure in brain mri images using binary operations, texture and set analysis. *IET Image Processing*, 11(9), 672–684.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.
- Pal, C., Das, P., Chakrabarti, A., & Ghosh, R. (2017). Rician noise removal in magnitude mri images using efficient anisotropic diffusion filtering. *International Journal of Imaging Systems and Technology*, 27(3), 248–264.
- Panattil, S. J., George, A., & Joy, M. M. (2021). The role of informational feedback as a game mechanic on user perceptions, attitudes and the intention to continue using a gamified health behaviour change support system. *Health Marketing Quarterly*, 39(1), 1–21.
- Panigrahi, L., Verma, K., & Singh, B. K. (2019). Ultrasound image segmentation using a novel multi-scale gaussian kernel fuzzy clustering and multi-scale vector field convolution. *Expert Systems with Applications*, 115, 486–498.
- Patel, A. P., Fisher, J. L., Nichols, E., Abd-Allah, F., Abdela, J., Abdelalim, A., Abraha, H. N., Agius, D., Alahdab, F., Alam, T., Allen, C. A., Anber, N. H., Awasthi, A., Badali, H., Belachew, A. B., Bijani, A., Bjørge, T., Carvalho, F., Catalá-López, F., ... Fitzmaurice, C. (2019). Global, regional, and national burden of brain and other cns cancer, 1990–2016: A systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, 18(4), 376–393.
- Pei, J., Chen, W., & Dong, X. (2017). Magnetic resonance imaging brain image segmentation method based on adaptive clustering algorithm. *Journal of Medical Imaging and Health Informatics*, 7(7), 1629–1635.
- Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging*, 35(5), 1240–1251.
- Pietka, E., Kawa, J., Badura, P., & Spinczyk, D. (2010). Open architecture computer-aided diagnosis system. *Expert Systems*, 27(1), 17–39.
- Portela, N. M., Cavalcanti, G. D., & Ren, T. I. (2014). Semi-supervised clustering for mr brain image segmentation. *Expert Systems with Applications*, 41(4), 1492–1497.
- Qin, J., Fu, W., Gao, H., & Zheng, W. X. (2016). Distributed k-means algorithm and fuzzy c-means algorithm for sensor networks based on multiagent consensus theory. *IEEE Transactions on Cybernetics*, 47(3), 772–783.

- Regnery, S., Knowles, B. R., Paech, D., Behl, N., Meissner, J.-E., Windisch, P., Ben Harrabi, S., Bernhardt, D., Schlemmer, H. P., Ladd, M. E., Rieken, S., Debus, J., & Adebeg, S. (2019). High-resolution flair mri at 7 tesla for treatment planning in glioblastoma patients. *Radiotherapy and Oncology*, 130, 180–184.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241).
- Saha, S., Alok, A. K., & Ekbal, A. (2016). Brain image segmentation using semi-supervised clustering. *Expert Systems with Applications*, 52, 50–63.
- Saladi, S., & Amutha Prabha, N. (2018). Mri brain segmentation in combination of clustering methods with markov random field. *International Journal of Imaging Systems and Technology*, 28(3), 207–216.
- Salem, S. A., Salem, N. M., & Nandi, A. K. (2009). Augmentation of a nearest neighbour clustering algorithm with a partial supervision strategy for biomedical data classification. *Expert Systems*, 26(1), 8–21.
- Scrucca, L. (2021). A fast and efficient modal em algorithm for gaussian mixtures. *Statistical Analysis and DataMining: The ASA Data Science Journal*.
- Sdiri, B., Beghdadi, A., Cheikh, F. A., Pedersen, M., & Elle, O. J. (2016). An adaptive contrast enhancement method for stereo endoscopic images combining binocular just noticeable difference model and depth information. *Electronic Imaging*, 2016(13), 1–7.
- Shanker, R., & Bhattacharya, M. (2019). Brain tumor segmentation of normal and lesion tissues using hybrid clustering and hierarchical centroid shape descriptor. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*.
- Siyal, M. Y., & Yu, L. (2005). An intelligent modified fuzzy c-means based algorithm for bias estimation and segmentation of brain mri. *Pattern Recognition Letters*, 26(13), 2052–2062.
- Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., ... Ye, X. (2017). Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mri. *International Journal of Computer Assisted Radiology and Surgery*, 12(2), 183–203.
- Tiwari, A., Srivastava, S., & Pant, M. (2020). Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. *Pattern Recognition Letters*, 131, 244–260.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310.
- Übeyli, E. D. (2008). Signal-to-noise ratios for measuring saliency of features extracted by eigenvector methods from ophthalmic arterial Doppler signals. *Expert Systems*, 25(5), 431–443.
- Von Luxburg, U. (2010). Clustering stability: An overview foundations and trends R. *Machine Learning*, 2(3), 235–274.
- Wadhwa, A., Bhardwaj, A., & Verma, V. S. (2019). A review on brain tumor segmentation of mri images. *Magnetic Resonance Imaging*, 61, 247–259.
- Wong, K.-P. (2005). Medical image segmentation: Methods and applications in functional imaging. In *Handbook of biomedical image analysis* (pp. 111–182). Springer.
- Wu, C., & Zhang, J. (2021). Robust semi-supervised spatial picture fuzzy clustering with local membership and kl-divergence for image segmentation. *International Journal of Machine Learning and Cybernetics*, 13(4), 1–25.
- Xu, J., Wang, N., & Wang, Y. (2018). Multi-pyramid image spatial structure based on coarse-to-fine pyramid and scale space. *CAA Transactions on Intelligence Technology*, 3(4), 228–234.
- Xue, Y., Xu, T., Zhang, H., Long, L. R., & Huang, X. (2018). Segan: Adversarial network with multi-scale l1 loss for medical image segmentation. *Neuroinformatics*, 16(3), 383–392.
- Yang, Y., Wu, C., Li, Y., & Zhang, S. (2020). Robust semisupervised kernelized fuzzy local information c-means clustering for image segmentation. *Mathematical Problems in Engineering*, 2020(3), 1–22.
- Yao, H., Duan, Q., Li, D., & Wang, J. (2013). An improved k-means clustering algorithm for fish image segmentation. *Mathematical and Computer Modelling*, 58(3–4), 790–798.
- Zappia, L., & Oshlack, A. (2018). Clustering trees: A visualization for evaluating clusterings at multiple resolutions. *GigaScience*, 7(7), giy083.
- Zarei Mahmoodabadi, S., Alirezaie, J., Babyn, P., Kassner, A., & Widjaja, E. (2013). A novel mcad for pediatric metabolic brain diseases incorporating dw imaging and mr spectroscopy. *Expert Systems*, 30(1), 21–33.
- Zhang, C., Shen, X., Cheng, H., & Qian, Q. (2019). Brain tumor segmentation based on hybrid clustering and morphological operations. *International Journal of Biomedical Imaging*, 2019, 1–11.
- Zhang, Y.-D., Zhao, G., Sun, J., Wu, X., Wang, Z.-H., Liu, H.-M., Govindaraj, V. V., Zhan, T., & Li, J. (2018). Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and jaya algorithm. *Multimedia Tools and Applications*, 77(17), 22629–22648.
- Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M. P., Oster, J., Wang, C., Smedby, Ö., Bian, C., Yang, X., Heng, P.-A., Mortazi, A., Bagci, U., Yang, G., Sun, C., Galisot, G., Ramel, J.-Y., Brouard, T., ... Yang, G. (2019). Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. *Medical Image Analysis*, 58, 101537.

AUTHOR BIOGRAPHIES

Michael Osadebey is currently an European Research consortium on Informatics and Mathematics (ERCIM) postdoctoral fellow at the Norwegian University of Science and Technology, Norway. In 2009, Michael obtained master degree with distinction in biomedical engineering from Tampere University of Technology, Finland. From October 2009 to December 2009, Michael was a Ragnar Granit research grant recipient. In 2015, Michael obtained his PhD in electrical and computer engineering from Concordia University, Montreal, Canada. Michael Osadebey, was until 2018, a MRI Reader at NeuroRx Research Inc., a Montreal-based clinical research organization (CRO). His duties at NeuroRx include the application of advanced image analysis software in the reading of MRI data of neurological disease patients undergoing clinical trial drug treatment.

Marius Pedersen is currently a professor at NTNU Gjøvik, Norway. He is also the director of the Norwegian Colour and Visual Computing Laboratory (Colourlab). His work is centered on image processing including subjective and objective image quality. Marius received his BSc

degree in computer engineering and MIT degree in media technology both from Gjøvik University College, Norway, in 2006 and 2007, respectively. He completed his PhD program in colour imaging from the University of Oslo, Norway, sponsored by Oce in 2011.

Meeta Kalra received the engineer degree from the Amity University, Noida, India in 2014 and was working as a Network Engineer until 2017 with Ericsson Global Private Ltd. and Amdocs, Gurgaon. She is currently pursuing her Masters degree with the Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, Quebec. Her research interests include information filtering and retrieval, data mining, and pattern recognition, computer vision, medical image segmentation and artificial intelligence.

Dag Waaler obtained PhD physics degree in 1983 from the Norwegian institute of Technology. He is an Associate Professor at the department of health sciences, Norwegian University of Science and Technology, Gjøvik, Norway and the head of the research group for welfare, innovation and technology at the same University. He was the head of department of Radiography and Health Technology at Gjøvik University College from 2005–2011. His research interests are medical imaging, image processing and image quality evaluation.

Nizar Bouguila received the engineer degree from the University of Tunis in 2000, the M.Sc. and Ph.D degrees from Sherbrooke University in 2002 and 2006, respectively, all in computer sciences. He is currently a full Professor within the Concordia Institute for Information Systems Engineering (CIISE) at Concordia University, Montreal, Quebec, Canada. His current research interests include computer vision and pattern recognition, machine learning and data mining, image and signal processing.

How to cite this article: Osadebey, M., Pedersen, M., Kalra, M., Waaler, D., & Bouguila, N. (2022). Enhancement of clustering techniques by coupling clustering tree and neural network: Application to brain tumour segmentation. *Expert Systems*, e13176. <https://doi.org/10.1111/exsy.13176>