 Acta Univ. Sapientiae Informatica **14,** 2 (2022) 316–337

DOI: 10.2478/ausi-2022-0018

A feature selection strategy using Markov clustering, for the optimization of brain tumor segmentation from MRI data

Ioan-Marius PISAK-LUKÁTS
Sapientia Hungarian University of Transylvania,
Cluj-Napoca, Romania
Óbuda University, Budapest, Hungary
Doctoral School of Applied Mathematics and
Applied Informatics
email: pisak.lukats@ms.sapientia.ro
ORCID:0000-0001-9518-6263

Levente KOVÁCS
Óbuda University, Budapest, Hungary
University Research, Innovation and Service Center
email: kovacs@uni-obuda.hu
ORCID:0000-0002-3188-0800

László SZILÁGYI
Computational Intelligence Research Group,
Sapientia Hungarian University of Transylvania,
Cluj-Napoca, Romania
Dept. of Electrical Engineering, Târgu Mureş
Óbuda University, Budapest, Hungary
University Research, Innovation and Service Center
email: lalo@ms.sapientia.ro
szilagyi.laszlo@uni-obuda.hu
ORCID:0000-0001-6722-2642

Computing Classification System 1998: I.5.3
Mathematics Subject Classification 2010: 68H10
Key words and phrases: machine learning, feature selection, Markov clustering, image segmentation

Abstract. The automatic segmentation of medical images stands at the basis of modern medical diagnosis, therapy planning and follow-up studies after interventions. The accuracy of the segmentation is a key element in assisting the work of the physician, but the efficiency of the process is also relevant. This paper introduces a feature selection strategy that attempts to define reduced feature sets for ensemble learning methods employed in brain tumor segmentation based on MRI data such a way that the segmentation outcome hardly suffers any damage. Initially, the full set of observed and generated features are deployed in ensemble training and prediction on testing data, which provide us information on all couples of features from the full feature set. The extracted pairwise data is fed to a Markov clustering (MCL) algorithm, which uses a graph structure to characterize the relation between features. MCL produces connected subgraphs that are totally separated from each other. The largest such subgraph defines the group of features which are selected for evaluation. The proposed technique is evaluated using the high-grade and low-grade tumor records of the training dataset of the BraTS 2019 challenge, in an ensemble learning framework relying on binary decision trees. The proposed method can reduce the set of features to 30% of its initial size without losing anything in terms of segmentation accuracy, significantly contributing to the efficiency of the segmentation process. A detailed comparison of the full set of 104 features and the reduced set of 41 features is provided, with special attention to highly discriminative and redundant features within the MRI data.

1 Introduction

According to recent estimations, 88,970 Americans would receive a primary brain tumor diagnosis in 2022, of which 18,280 would die from a malignant brain tumor [31]. The advantages of the use of image segmentation of brain tumors cannot be overemphasized, given the non-intrusive nature of the imaging process and that early detection increases the chances of survival [23].

The automatic segmentation of brain tumors from MRI data has been an intensively researched topic for at least thirty years now [5, 16]. The established solutions grew together with the development of computers from various points of view. Early solutions could only handle single cross sections of the brain, using a single weighting scheme of the MRI device. Later the processing of volumetric data and multi-spectral volumes became possible, as the available storage space continuously grew year by year. With the development of CPUs and GPUs, more and more complex algorithms emerged, which led to better segmentation outcome.

The MICCAI Brain Tumor Segmentation Challenge (BraTS) [15], organized yearly in the last decade, provoked an intensification of efforts in brain tumor segmentation algorithm development. In the early years of BraTS, the whole arsenal of classical machine learning algorithms was deployed in various combinations. The best solutions in this era were obtained by the use of AdaBoost classifier [8], random forests [11, 12], extremely random trees [21], and other ensembles [20]. On the other hand, in the recent years of the BraTS challenges [2], convolutional neural networks (CNN) and deep learning outweighed the classical solutions and provided better segmentation accuracy, by the use of much more complex network models [9, 13]. The U-net architecture introduced by Ronneberger et al. [24] is employed in brain tumor segmentation in many variants and combinations. Very recent CNN architectures also include special attention mechanisms [4, 14] to optimize the segmentation outcome.

One of the biggest differences between classical machine learning based and CNN based solutions is the way of using the features. CNN based architectures have their convolutional layers which compute the features used for classification from the original set of images collected by the MRI device [3]. On the other hand, classical machine learning needs handcrafted features, which are computed for each pixel of each volume, usually separately in each data channel independently of each other. There exist ensemble learning based methods in the literature that use hundreds of handcrafted features together with the four originally observed channels. Whenever a test record is presented to the trained ensemble for segmentation, all these hundreds of features need to be computed for each brain pixel in the volume, which may last a prohibitively long time. This is why, for such cases it is useful to perform a feature selection, to establish a reduced set of features which can produce practically the same segmentation accuracy, but can be computed more efficiently. The role of this feature selection is to eliminate the redundant and noisy features, which do not add relevant information to the feature vector and thus represent obstacles to the effective and efficient classification of pixels.

In a previous study of our research group, Győrfi et al [7] introduced a strategy for feature ranking and selection, depending on how frequently each feature is employed in making decisions regarding the pixels of test records, and how often the decisions to which each feature contributed is correct. In our previous paper [22] we extended the feature monitoring to couples of features, extracting information on how often all couples of features take part in making decisions regarding individual test pixels, and further on, what part of these decisions is correct. In this paper, we employ a Markov clustering algorithm to extract groups of features which are likely to work effectively

and make correct decisions together, based on the above mentioned pairwise observations. The obtained reduced feature sets are evaluated on high-grade and low-grade tumor records of the BraTS 2019 training datasets, and their performance is compared with various combinations of the generated feature groups.

The rest of this paper is structured as follows: Section 2 presents the proposed Markov clustering based feature selection scheme. Section 3 provides a detailed evaluation of the proposed feature selection algorithm via deploying it in brain tumor segmentation using the training dataset of the BraTS 2019 challenge. Section 4 discusses the main findings of the numerical evaluation process, while Section 5 concludes the study.

2 Methods

2.1 Problem formulation

There is a brain tumor segmentation solution built by our team five years ago [26] and further developed recently [6], which uses ensemble learning based on binary decision trees [1]. The ensemble can be trained with any dataset originating from the MICCAI Brain Tumor Segmentation Challenge (BraTS) [15, 2]. In our current study we use the training dataset of the BraTS 2019 challenge to train and evaluate the segmentation performance of the ensemble. Beside the four observed data channels (T1, T2, T1C and FLAIR), we generate 100 further handcrafted features for each brain pixel of the volumes. The existing system can produce segmentation accuracy characterised by average Dice scores around 84%. Our main goal is to reduce the feature vector size as much as possible without losing significantly from the segmentation quality. The aim is to define some reduced feature sets which can be extracted with less computations, and maintain the best segmentation quality that the ensemble can offer.

2.2 Data

The training dataset published by the BraTS 2019 challenge [15, 2] contains $n_{\mathrm{LGG}} = 76$ low-grade glioma (LGG) records and $n_{\mathrm{HGG}} = 259$ high-grade glioma (HGG) records. Each record consists of 1.2-1.6 million volumetric pixels, each representing $1\,\mathrm{mm}^3$ of brain tissues. Non-brain tissues are removed from the volumes. All volumes were recorded with four different weighting schemes frequently used in MRI imaging. The data collected in the four data

channels (T1, T2, T1C and FLAIR) were registered together. An expert made label is assigned to each pixel of the records that can be used as expected value at training and ground truth at validation and prediction. Most volumes have pixels with missing data, meaning that in one or more data channels the observed value is zero, which is an invalid value. Some volumes only have some single, isolated pixels with missing data, which can be easily filled with average values extracted from neighbor pixels. Other records contain large patches with pixels having missing data, frequently in the most important data channel called FLAIR. These low-quality records were intentionally inserted into the BraTS dataset to see, which segmentation algorithms can effectively handle such odd cases. Details regarding missing data in the BraTS dataset are reported in [6].

2.3 Procedure

In this section we give a brief description of the whole MRI data processing procedure. Further details can be found in our previous work [26]. The contribution of this study focuses on the feature selection step, which is described in Section 2.4.

MRI data is fed to a four-stage processing procedure:

1. Preprocessing aims at the elimination of intensity non-uniformity [32, 29] in the individual MRI volumes, which is performed by the N3 algorithm proposed by Sled et al. [25], followed by filling the missing data with averaged values from the neighborhood where these are available, or with global average extracted from the volume in question. The next preprocessing step consists in histogram normalization or uniformization, which is necessary because the MRI devices do not have a standardized intensity scale, the numerical values have to be interpreted within their context. This histogram transformation is performed on each of the four data channels separately, using the algorithm proposed by Nyúl et al. [19], following the recommendations given in [10] regarding the reduced number of constraints.

2. The second processing step consists in feature generation, which is motivated by three facts at least:

   - The correlation between the intensity of neighbor pixels is strong, so the neighborhood may help reduce the level of high-frequency noise, and can give additional information for the description of the pixel in question.

- The four data channels of an MRI record are never perfectly registered to each other, so the information on a certain piece of brain tissue can be situated at coordinates (x, y, z) in a certain data channel, and at $(x \pm dx, y \pm dy, z \pm dz)$ in another data channels, where dx, dy, and dz represent short displacements.
- There can be additional information in the textures as well, which motivates using features extracted by methods like wavelet transform [18].

Here we intend to add an extra feature selection step to the image data processing procedure, which is presented in Section 2.5.

3. The third processing stage consists in the classification of pixels, which is performed by an ensemble of binary decision trees, as described in our previous work [26]. Whichever dataset we are using, those with HGG or LGG tumors, the whole set of records is randomly divided into two equal or almost equal groups, let us denote them by G_A and G_B. When G_A is used to train the ensemble, the ensemble is used to perform the prediction on G_B, and vice versa. All ensembles consist of 75 decision trees, which are not limited in depth during training. The training process uses 100 thousand randomly selected feature vectors from each record of the training dataset, out of which 92% are negatives and 8% positives according to the ground truth. Each node of the decision trees is trained using the standard entropy-based rule [1]. Leaves of the decision trees are those nodes which have nothing to separate, they only need to handle either positive or negative labeled pixels. When used for prediction, each decision tree of the ensemble gives a dedicated vote to each pixel of the test record. The decision referring to a given pixel is established on the majority of votes it receives. This label is not yet the final one.

4. The decision made by the ensemble regarding any of the test pixels is examined again in a post-processing stage, and overruled whenever necessary. A morphological criterion is used to perform this step. The rate of pixels declared positive by the ensemble is established within a spatial neighborhood of the pixels ($11 \times 11 \times 11$ and $9 \times 9 \times 9$ sized neighborhood in case of LGG and HGG records, respectively) and compared with a predefined threshold 1/3, which was established empirically. The final decision regarding any pixel is positive, if the rate of positives exceeds this threshold, and negative otherwise, no matter what the initial

Feature set	Operation	Neighborhood	Data channels				Total features
			T1	T2	T1C	FLAIR	
OBS	none	1×1	1	1	1	1	4
BASE	average	$3 \times 3 \times 3$	1	1	1	1	12
	maximum	$3 \times 3 \times 3$	1	1	1	1	
	minimum	$3 \times 3 \times 3$	1	1	1	1	
AVG	average	3×3	1	1	1	1	20
	average	5×5	1	1	1	1	
	average	7×7	1	1	1	1	
	average	9×9	1	1	1	1	
	average	11×11	1	1	1	1	
GRAD	gradient in four main directions	7×7	4	4	4	4	16
MED	average	3×3	1	1	1	1	20
	average	5×5	1	1	1	1	
	average	7×7	1	1	1	1	
	average	9×9	1	1	1	1	
	average	11×11	1	1	1	1	
GABOR	Gabor wavelet in eight main directions	15×15	8	8	8	8	32
Total features in full set of features			26	26	26	26	104

Table 1: Inventory of features in the full feature set

decision was. Previous works of our research group [26, 6] found this post-processing step beneficial to segmentation accuracy.

2.4 Feature generation

We previously stated that the multi-spectral image data collected by the MRI device can offer much more support for accurate segmentation than the four observed feature values of individual pixels. Consequently, we defined 100 further handcrafted features, 25 from each data channel, that were extracted from the image data considering various neighborhoods of each brain pixel. The full set of features is presented in Table 1, and it consists of the following feature groups:

1. Feature group OBS consists of the four observed intensity values of each pixel.

2. Feature group BASE contains maximum, minimum and average values extracted from $3 \times 3 \times 3$ sized spatial neighborhood of the pixel.

3. Feature groups AVG and MED consist of average and median values, respectively, extracted from planar neighborhoods of size $k \times k$ with $k = 3, 5, \ldots 11$.

4. Feature group GRAD contains gradient values extracted from 7×7 neighborhood, with respect to four main directions, according to the method presented in [6].

5. Feature group GABOR contains eight values extracted from each data channel, with respect to eight main directions, using neighborhoods of size 15×15. The OpenCV implementation of Gabor wavelets were employed to establish these feature values [17].

Each of the computed features used values taken from only one data channel. Neighbor pixels that do not represent brain were neglected in all cases.

2.5 Feature selection

Whenever a vote is cast by a decision tree of the ensemble, concerning a given pixel of an MRI record from the testing dataset, a certain route is followed from the root of the tree to one of its leaves, based on the feature vector of the test pixel. Each node along the route compares the value of a single feature of the test vector with a threshold, and then the next node depends on the outcome of the comparison. This way, while we follow the route toward a leaf, we use a subset of features Ψ_0 to make the decision, which is either correct or not. In this study, we are attempting to characterise the relation between any couple of features $f_i, f_j \in \Psi$, where Ψ represents the full set of features, by two measures:

1. What is the ratio of decisions whose subset of features Ψ_0 contains both features f_i, f_j? We denote this value by φ_{ij}.

2. What is the ratio of correct decisions out of those decisions whose subset of features Ψ_0 contains both features f_i, f_j? We denote this value by κ_{ij}.

Obviously both φ_{ij} and κ_{ij} are in the interval $[0, 1]$. The φ values tend to have a distribution with high variance: there are few couples of features which appear in most decisions, while most couples of features appear in only a few percent of the decisions. On the other hand, the most part of κ_{ij} values are situated between 0.6 and 0.9, here the variance is much lower. In case of $\varphi_{ij} = 0$, which does happen in very few cases, we also set $\kappa_{ij} = 0$ instead of dividing by zero.

We are creating clusters of features based on the pairwise relations between features. This is performed by the use of Markov clustering (MCL), which is a graph-based algorithm relying on pairwise similarity measures of the input data. In our case, the input data is represented by the $|\Psi| = 104$ features, while the starting similarity matrix is derived from the φ_{ij} and κ_{ij} values. Due to the distributions described above, we found most suitable to define the input pairwise similarity matrix $\mathbf{S} = \{s_{ij}\}_{i,j=1\ldots|\Psi|}$ with values established as $s_{ij} = \varphi_{ij} \cdot \kappa_{ij}^{\alpha}$, where $\alpha = 3$ was set empirically.

Our Markov clustering implementation is an iterative algorithm that performs the following operations in every cycle:

- *Inflation* favors higher similarity values and suppresses lower ones. It is performed according to the formula $s_{ij} \leftarrow s_{ij}^{r}\ \forall i, j = 1 \ldots |\Psi|$, where $r > 1$ represents the inflation rate, which is a predefined constant. The rows or the columns are normalized after inflation.

- *Symmetrization* is a necessary step to keep the matrix close to symmetric. It is accomplished by the formula

$$s_{ij}, s_{ji} \leftarrow \begin{cases} \sqrt{s_{ij} s_{ji}} & \text{if } \sqrt{s_{ij} s_{ji}} > \varepsilon \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

  where $i = 1 \ldots |\Psi|$, $j = i+1 \ldots |\Psi|$, and ε is a predefined threshold, in our case $\varepsilon = 0.002$. This formula also removes weak connections from the graph. The rows or the columns are normalized after symmetrization.

- *Expansion* is an operation that favors longer walks in the graph represented by the similarity matrix $\mathbf{S}$, and is accomplished via matrix squaring: $\mathbf{S} \leftarrow \mathbf{S}^2$.

Inflation and expansion work against each other, while the only main parameter regulating this phenomenon is the inflation rate. Inflation tends to tear the graph into isolated subgraphs, while expansion tends to fill the existing subgraphs with complete set of edges. The algorithm converges in a few

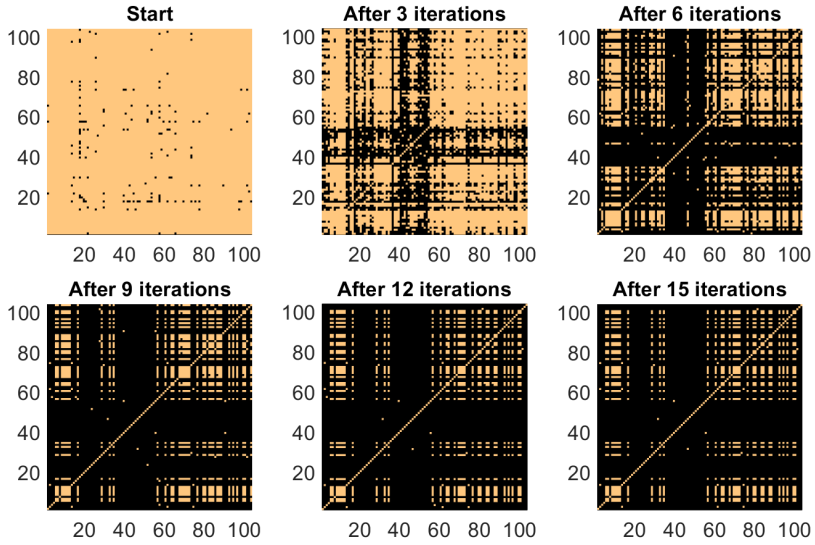


Figure 1: The evolution of the similarity matrix during the MCL iterations. Nonzero values are shown in orange, zeros appear in black. The input matrix is obtained from a full test on LGG records. The inflation rate is $r = 1.55$.

dozen of cycles, after which there are several isolated subgpraphs in the graph, whose submatrix is filled with almost equal values. Due to the nature of the MCL algorithm and our input data, a single large cluster of nodes is formed within the graph, while all other nodes will be separated into singletons or tiny groups. We will call reduced set of features the collection of features corresponding to the nodes that stayed within the largest connected group of nodes. The inflation rate strongly influences the size of the reduced feature set.

Markov clustering algorithm has several versions with computations optimized against the clock [27, 30, 28], but in our study we did not need them because clustering a graph of 104 nodes is performed in less then a second.

Figure 1 presents the evolution of the Markov similarity graph during the iterative clustering process. The algorithm seems to converge in 12 iterations.

2.6 Evaluation

This study employs statistical indicators to evaluate the segmentation accuracy achieved by the use of various feature subsets. These indicators are extracted from the confusion matrix. In case of two-class separation, the con-

fusion matrix contains four values, namely the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) pixels. From the confusion matrix of each test, we extract the following statistical accuracy indicators:

1. sensitivity or recall or true positive rate, defined as

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ , \tag{2}$$

2. specificity or true negative rate, defined as

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \ , \tag{3}$$

3. positive predictive value or precision, defined as

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \ , \tag{4}$$

4. Dice similarity coefficient or Dice score or F1-score, defined as

$$\mathrm{DSC} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \ , \tag{5}$$

5. correct decision rate or accuracy, defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \ . \tag{6}$$

These values are extracted separately from each MRI record used for testing, and can be used to establish global accuracy indicators like average, median, and various percentile values. Our study uses the Q1, Q2, Q3 quartiles, and the average of the values falling between Q1 and Q3. The motivation of this choice is given in Section 3.

Dice score and accuracy are strongly correlated, but due to the imbalance of positive and negative pixels, DSC values vary in a wider range than ACC values, and thus the Dice score is more suitable to express the quality difference between two segmentation outcomes.

All these statistical indicators are positive real variables with values up to 1. Higher values always indicate better segmentation accuracy.

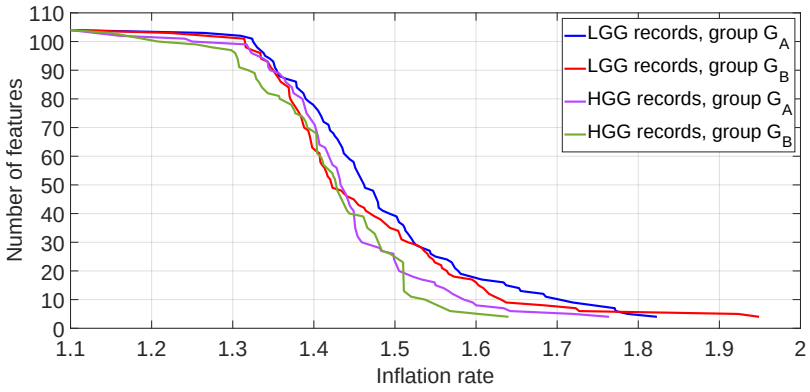


Figure 2: The size of the largest cluster of features produced by the Markov clustering algorithm, plotted against the inflation rate $\mathtt{r}$.

3 Results

The proposed methodology was applied separately to the 259 HGG tumor records and the 76 LGG tumor records of the BraTS 2019 dataset. The full set of 76 or 259 records was randomly divided into two (almost) equal groups denote by $\mathtt{G_A}$ and $\mathtt{G_B}$. These two groups took turns in serving as training data and testing data for the ensemble of the binary decision trees. Initially the training and testing employed the full set of 104 features. The feature similarity matrix obtained from the test run on group $\mathtt{G_A}$ was later used to define reduced feature sets for the tests of $\mathtt{G_B}$, and vice versa. Consequently we obtained four series of reduced feature sets, two for LGG records and two for HGG records. If we deploy MCL with a sufficiently low inflation rate $\mathtt{r}_{\min}$, all 104 features remain connected in the similarity graph. By increasing the inflation rate in small steps, it is possible to get reduced sets of features with small decrements of the feature count: from 104 features we get down to 5 features in up to 70 steps. For almost any number of features between 5 and 104, there exists an inflation rate that can produce a largest group of features with the desired size. Figure 2 presents the evolution of the reduced feature set size plotted against the inflation rate for the four series of reduced feature sets.

Beside the tests performed using the reduced sets of features obtained via Markov clustering, further training and testing was performed to evaluate the

performance of various subsets obtained via combining certain feature groups, as follows:

- All combinations of the four data channels were encoded into feature subsets, which contained only one, two, or three data channels out of the four available ones, resulting in subsets of 26, 52, or 78 features. These tests can help us establish the significance of each data channel in terms of segmentation quality.

- The six feature groups presented in Table 1, can also form feature subsets. Here we formed all combination of BASE, AVG, GRAD, MED, and GABOR feature groups and added to all combinations the OBS features. This way we obtained feature subsets with size varying between 16 and 92. Tests running on these feature sets can give us insight into the role of these types of features in the segmentation process.

All the above listed subsets of features can serve as comparison for the subsets generated by MCL. Our subsets of features suggested by MCL are suitable for high quality segmentation, if they outperform the combinations of feature groups having the same number of features.

In case of any test, no matter whether it employed HGG or LGG data, or it was using record group G_A or G_B for testing, or which subset of features was deployed, we established the Dice scores for each MRI record separately, and from these values we extracted the following indicators: the Q1, Q2 and Q3 quartiles (or in other words, the 25, 50, and 75 percentiles) of the Dice score values, and the average of those falling between Q1 and Q3. The usage of these indicators is motivated by the fact that for us it is not the most important to raise the quality of the worst segmentation outcomes because those values can change the most. A segmentation should not be declared that best one just because it can give 30% Dice score on a record where others produce only 10%. Better performance should be judged by the mass of good-quality MRI records, where the Dice score is above 75%. The question is, which feature subset performs best in this domain.

In this order, Figures 3-6 present Q1, Q2, Q3, and average of those between Q1 and Q3 values of the Dice scores obtained during the four series of tests. Figures 3 and 4 exhibit the segmentation quality benchmarks of tests run on HGG records, while Figures 5 and 6 those run on LGG records. Further on, Figures 3 and 5 exhibit the segmentation quality benchmarks of tests run on record group G_A from HGG and LGG datasets, respectively, while Figures 4 and 6 those run on record group G_B from HGG and LGG datasets, respectively.

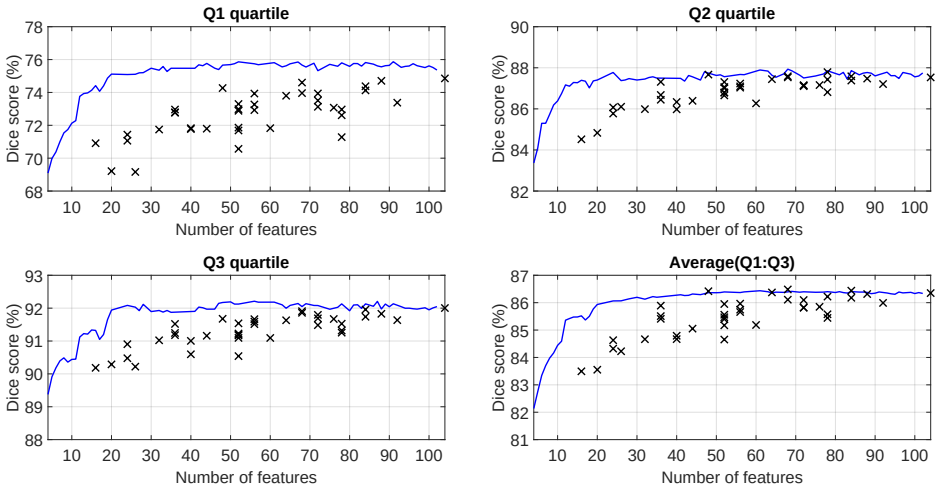


Figure 3: Global Dice score based benchmarks (Q1, Q2, Q3 quartiles and average of values between Q1 and Q3), achieved by various reduced feature sets (blue graph) and combinations of feature groups (black crosses) on group G_A of HGG tumor records, plotted against the number of features.

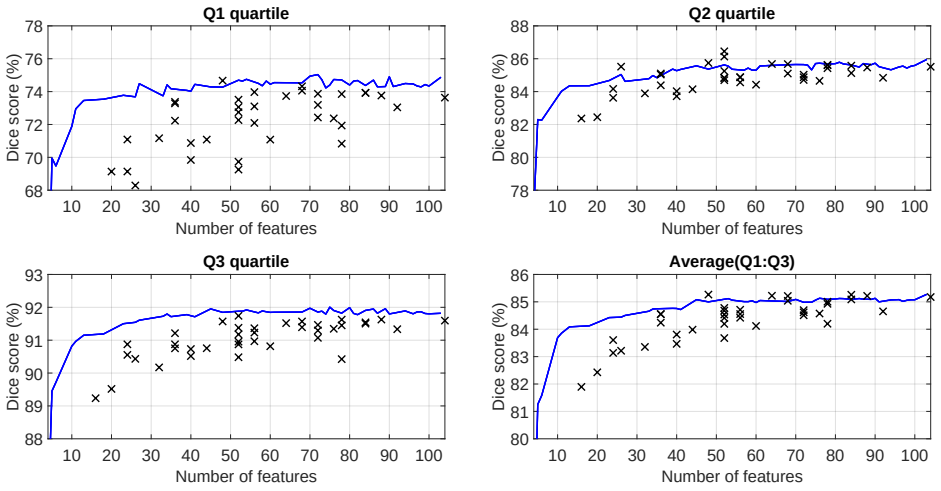


Figure 4: Global Dice score based benchmarks (Q1, Q2, Q3 quartiles and average of values between Q1 and Q3), achieved by various reduced feature sets (blue graph) and combinations of feature groups (black crosses) on group G_B of HGG tumor records, plotted against the number of features.

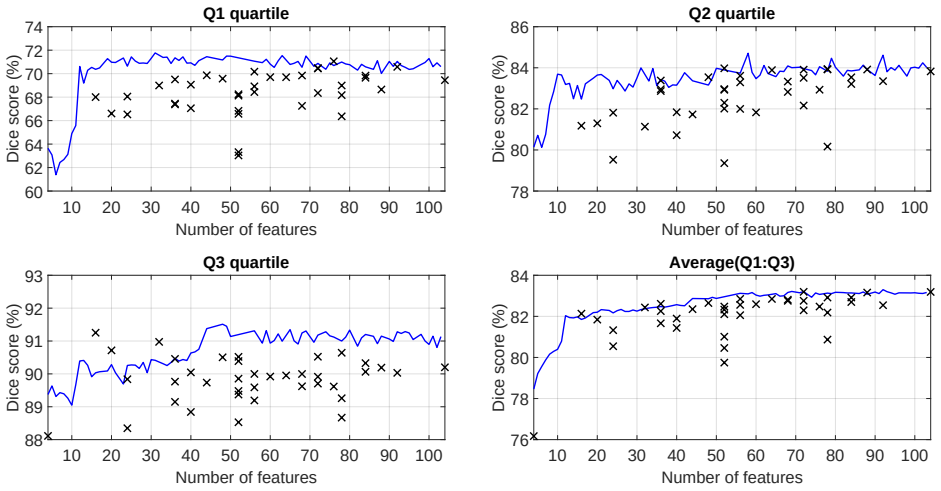


Figure 5: Global Dice score based benchmarks (Q1, Q2, Q3 quartiles and average of values between Q1 and Q3), achieved by various reduced feature sets (blue graph) and combinations of feature groups (black crosses) on group G_A of LGG tumor records, plotted against the number of features.

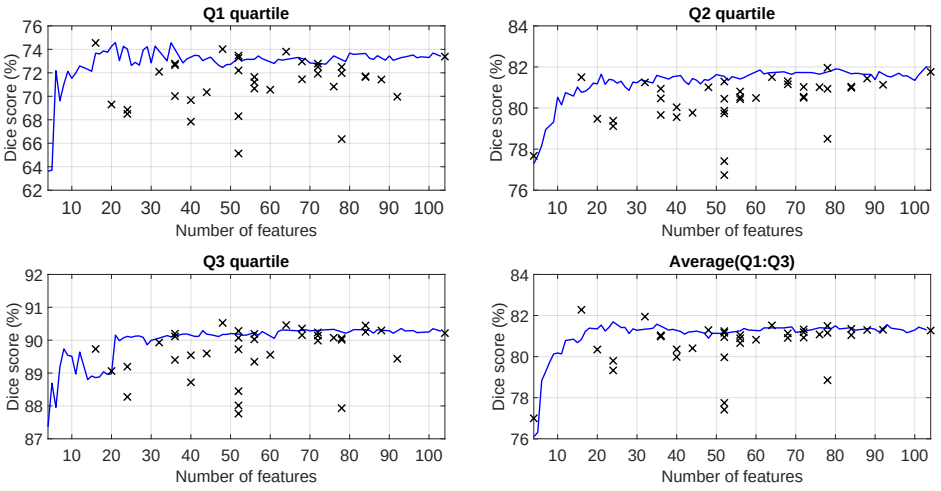


Figure 6: Global Dice score based benchmarks (Q1, Q2, Q3 quartiles and average of values between Q1 and Q3), achieved by various reduced feature sets (blue graph) and combinations of feature groups (black crosses) on group G_B of HGG tumor records, plotted against the number of features.

Feature group	Appearances					Percent
	0	1	2	3	4	
Observed	3	-	1	-	-	12.50%
Spatial	2	2	3	2	3	54.17%
Average	10	4	2	3	1	26.25%
Gradient	11	3	2	-	-	10.94%
Median	9	2	3	2	4	37.50%
Gabor wavelet	2	5	9	9	7	60.94%
T1 features	13	3	4	3	3	30.77%
T2 features	13	6	4	2	1	23.08%
T1C features	10	3	6	3	4	38.46%
FLAIR features	1	4	6	8	7	65.38%

Table 2: Average structure of a 41-feature reduced feature set

In each of these figures we can see four panels, one for each quartile or average of Dice scores. In each of these panels we can see a continuous blue line, which indicates the Dice score benchmark value obtained by MCL-made subsets of features, plotted against the number of features in the subset. The black crosses in each panel indicate the Dice score benchmark value obtained by the counter-candidate subsets of features formed by various groups of features (combination of data channels or combination of BASE, AVG, MED, GRAD and GABOR), also plotted against the number of features in the subset in question. The blue lines in these graphical representations indicate that in most cases we can reduce the number of features from 104 to a value between 30 and 40 without losing anything from the segmentation quality reflected in Dice scores. The position of the black crosses ($\times$), whose majority is situated below the blue graphs suggest that the subsets of features given by MCL produce better segmentation quality than most combinations of feature groups available in the full set of features.

Let us examine the specific case of reduced features sets containing 41 features. We chose this number such a way that is situated at the lower end of the range where the Dice scores obtained by the full feature set are preserved. Another reason of this choice is the fact that in all four test series we found a reduced feature set with this size. We examined the features in these four reduced feature sets, the findings are exhibited in Table 2. This table indicates how many items of the groups of features involved in this study are present in these reduced feature sets. For any group of features we can see how many of

Accuracy indicator	Data-set	Q1		Q2		Q3		Avg(Q1:Q3)	
		Ψ_{41}	Ψ	Ψ_{41}	Ψ	Ψ_{41}	Ψ	Ψ_{41}	Ψ
Recall	HGG	78.69	79.07	91.89	91.85	96.00	96.04	90.07	90.17
(TPR)	LGG	68.03	67.93	85.42	85.63	92.55	92.04	83.52	83.28
Precision	HGG	81.06	80.88	87.12	86.95	93.09	92.86	87.17	87.01
(PPV)	LGG	77.40	75.78	90.01	89.35	95.06	94.50	88.67	88.42
Accuracy	HGG	97.89	97.87	98.47	98.46	99.00	99.00	98.46	98.46
(ACC)	LGG	96.56	96.64	98.11	98.15	98.78	98.81	97.90	97.88
Dice score	HGG	77.75	77.98	87.42	87.57	90.72	90.76	86.51	86.58
(DSC)	LGG	74.29	73.38	81.97	81.83	89.71	90.20	82.19	82.11

Table 3: Statistical accuracy indicators: 41-feature reduced feature set (Ψ_{41}) vs. full set of features (Ψ)

its items appear in 0, 1, 2, 3, and 4 of these four reduced feature sets, respectively. The final column indicates the percentage of the features within the feature group that is present in the four reduced feature sets of 41 elements each. Within this column 100% would mean that all elements of a feature group are present in all four reduced feature sets. Of course, this is not the case for any of the feature groups, but we can read from this table, that Gabor wavelet features are used most in these reduced sets, and FLAIR is the most relevant data channel in brain tumor segmentation. Table 2 also indicates that 15 features out of the 104 ones of the full set of features are present in all four subsets Ψ_{41}. Further 16 features are present in three of the four feature sets. Half of the Gabor wavelet features are among the most important ones, while none of the gradient or the original observed features appear to be relevant.

Now let us compare the accuracy we can achieve using the full set of features (Ψ) and with the reduced set of 41 features (Ψ_{41}). Table 3 presents the main accuracy benchmark indicators listed in Section 2.6 for both HGG and LGG data. For each statistical indicator, the Q1, Q2 and Q3 quartiles, and the average between Q1 and Q3 are reported. Numerical values are highlighted with various shades of green and red, green representing the higher and red the lower accuracy. Here we merged the two runs on HGG records and the other two on LGG records. This was possible because we had reduced feature sets of 41 items in all four runs. Feature set Ψ_{41} seems to be slightly better in precision, while the full feature set has better Dice score, but there is no remarkable difference between the two performances.

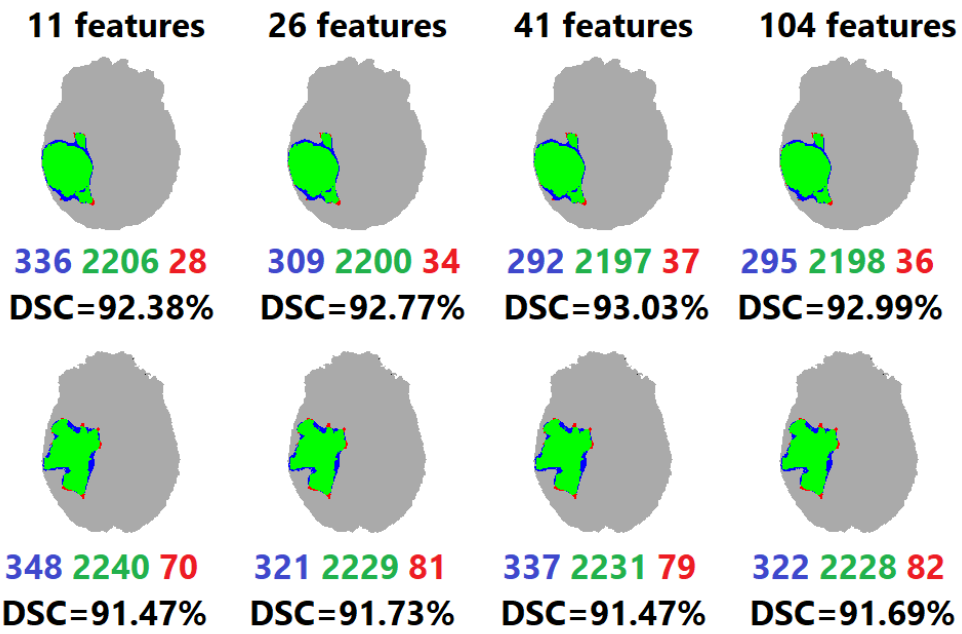


Figure 7: Segmentation results demonstrated on two cross sections, in case of 11, 26, 41, and 104 features. Green color represents true positives, blue indicates false positves, and red color shows false negatives. Blue, green, and red numbers under each slice shows the number of pixels painted with the respective color. Dice scores were computed for the represented slice, not the whole volume.

Figure 7 shows two selected cross sections and their segmentation results using all 104 features, and three reduced features sets of 11, 26, and 41 features. While there is no visible quality difference among the slices shown in any of the both rows, the numerical benchmarks indicate small differences. While it is hard to define, which reduced set of features is best, we can remark that reduced feature sets can successfully compete with the full set of features.

4 Discussion

The proposed feature reduction technique was evaluated on four sets of data, two groups of records G_A and G_B from both HGG and LGG tumor datasets of BraTS 2019. Each of this four series of tests used the pairwise feature con-

tribution data of the complementary group (tests on G_A used feature data of G_B, and vice versa). Detailed accuracy benchmarks exhibited in Figures 3-6 revealed that carefully selected 30% of the full set of features can provide segmentation quality comparable to the original. When the reduction goes beyond a certain limit, which is situated approximately at 25% of the initial size of the feature set, segmentation accuracy benchmarks drastically drop, suggesting that this low number of features does not contain enough information for an optimal segmentation.

The latter observation suggests that segmentation methods based on convolutional neural networks, which usually only receive observed data (cross sections in case of 2D convolution and whole volumes in case of 3D convolution), may work better if we replace some of the data channels with some handcrafted features organized in the same format.

The only relevant parameter of the proposed feature selection method is the inflation rate of the Markov clustering algorithm. As Figure 2 shows, there is no direct formula to establish the inflation rate that is suitable to produce a reduced feature set of a desired size. For example, there are significant differences in the inflation rate used to obtain 41 features in four different data subsets. If we wish to establish a reduced set of features that contains approximately k features, we may turn to the interval halving strategy. There is a lower limit, e.g. $r_{lo} = 1.1$ which keeps together the whole set of features, and an upper limit, e.g. $r_{lo} = 2.0$ which leads to a collection of single nodes in the Markovian similarity graph. Halving the interval 10 to 20 times according to the outcome of Markov clustering with inflation rate $(r_{lo} + r_{hi})/2$, we can obtain the feature subset of approximately the desired cardinality.

The reduced feature set that leads to the same segmentation quality as the full initial feature set can support the optimization of the brain tumor segmentation process. Having to compute less feature data when the trained ensemble receives a new record to segment causes substantial reduction in both execution time and required storage space.

5 Conclusion

In this paper we introduced a feature selection strategy based on Markov clustering, which relies on the evaluation of the contribution of all possible feature pairs (f_i, f_j) within the initial whole feature set Ψ. The proposed strategy was employed in a brain tumor segmentation problem involving both HGG and LGG tumor records. The proposed method apparently can select one third of

the features and maintain the segmentation quality that can be obtained with the full set of features. Such a reduction in the number of features can save a lot of storage space and execution time in any image segmentation framework.

Acknowledgements

This study was supported in part by the Institute for Research Programs of Sapientia University and the Consolidator Researcher Program of Óbuda University. The authors are grateful to Csaba Sámuel Nagy for his efforts in finding Gabor wavelets that effectively support the brain tumor segmentation process.

References

[1] S. B. Akers, Binary decision diagrams, *IEEE Transactions on Computers* **C-27** (1978) 509–516. ⇒319, 321

[2] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, M. Prastawa, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, arXiv (2019) 1181.02629v2. ⇒318, 319

[3] S.R. Chandaran, G. Muthusamy, L.R. Sevalaiappan, N. Senthilkumaran, Deep learning-based transfer learning model in diagnosis of diseases with brain magnetic resonance imaging, *Acta Polytechnica Hungarica* **19**, 5 (2022) 127–147. ⇒318

[4] Y.K. Chang, Z.Z. Zheng, Y.W. Sun, M.M. Zhao, Y. Lu, Y. Zhang, DPAFNet: A residual dual-path attention-fusion convolutional neural network for multimodal brain tumor segmentation *Biomedical Signal Processing and Control* **79** (2023) 104037. ⇒318

[5] N. Gordillo, E. Montseny, P. Sobrevilla, State of the art survey on MRI brain tumor segmentation, *Magnetic Resonance Imaging* **31** (2013) 1426–1438. ⇒317

[6] Á. Győrfi, L. Szilágyi, L. Kovács, A fully automatic procedure for brain tumor segmentation from multi-spectral MRI records using ensemble learning and atlas-based data enhancement, *Applied Sciences* **11** (2021) 564. ⇒319, 320, 322, 323

[7] Á. Győrfi, L. Kovács, L. Szilágyi, A feature ranking and selection algorithm for brain tumor segmentation in multi-spectral magnetic resonance image data, *Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)* (2019) 804–807. ⇒318

[8] A. Islam, S. M. S. Reza, K. M. Iftekharuddin, Multifractal texture estimation for detection and segmentation of brain tumors, *IEEE Transactions on Biomedical Engineering* **60** (2013) 3204–3215. ⇒318

[9] K. Kamnitsas, C. Ledig, V.F.J. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, *Medical Image Analysis* **36** (2017), 61–78. ⇒318

[10] A. Kőble, Á. Győrfi, Sz. Csaholczi, B. Surányi, L. Dénes-Fazakas, L. Kovács, L. Szilágyi, Identifying the most suitable histogram normalization technique for machine learning based segmentation of multispectral brain MRI data, *IEEE AFRICON* (2021) 71–76. ⇒320

[11] L. Lefkovits, S. Lefkovits, L. Szilágyi, Brain tumor segmentation with optimized random forest, *International MICCAI Brainlesion Workshop*, *Lecture Notes in Computer Science* **10154** (2017) 88–99. ⇒318

[12] S. Lefkovits, L. Lefkovits, L. Szilágyi, Applications of different CNN architectures for palm vein identification, *Modeling Decisions for Artificial Intelligence (MDAI)*, *Lecture Notes in Computer Science* **11676** (2019) 295-306. ⇒318

[13] S. Lefkovits, L. Lefkovits, L. Szilágyi, HGG and LGG brain tumor segmentation in multi-modal MRI using pretrained convolutional neural networks of Amazon Sagemaker, *Applied Sciences* **12** (2022) 3620. ⇒318

[14] H.X. Liu, G.Q. Huo, Q. Li, X. Guan, M.L. Tseng, Multiscale lightweight 3D segmentation algorithm with attention mechanism: Brain tumor image segmentation, *Expert Systems with Applications* **214** (2023) 119166. ⇒318

[15] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The multimodal brain tumor image segmentation benchmark (BRATS), *IEEE Transactions on Medical Imaging* **34**, 10 (2015) 1993–2024. ⇒318, 319

[16] G. Mohan, M.M. Subashini, MRI based medical image analysis: Survey on brain tumor grade classification, *Biomedical Signal Processing and Control* **39** (2018) 139-161. ⇒317

[17] Cs.S. Nagy, Optimization of a tumor recognition algorithm (Hungarian), MSc thesis, Sapientia University, Romania, 2022. ⇒323

[18] T.Ö. Onur, Improved image denoising using wavelet edge detection based on Otsu's thresholding, *Acta Polytechnica Hungarica* **19**, 2 (2022) 79–92. ⇒321

[19] L. G. Nyúl, J. K. Udupa, X. Zhang, New variants of a method of MRI scale standardization, *IEEE Trans. Med. Imaging* **19**, 2 (2010) 143–150. ⇒320

[20] A. Phophalia, P. Maji, Multimodal brain tumor segmentation using ensemble of forest method, *Proc. 3rd International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*(BraTS MICCAI 2017, Quebec City), *Lecture Notes in Computer Science* **10670** (2018) 159–168. ⇒318

[21] A. Pinto, S. Pereira, D. Rasteiro, C. A. Silva, Hierarchical brain tumour segmentation using extremely randomized trees, *Pattern Recognition* **82** (2018) 105–117. ⇒318

[22] J.M. Pisak-Lukáts, L. Szilágyi, Markov clustering based feature selection for brain tumor segmentation from multi-spectral MRI records, *20th Jubilee IEEE World Symposium on Applied Machine Intelligence and Informatics (SAMI)* (2022) 165–170. ⇒318

[23] R. Ranjbarzadeh, A. Caputo, E.B. Tirkolaee, S.J. Ghoushchi, M. Bendechache, Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools, *Computers in Biology and Medicine* **152** (2023) 106405. ⇒317

[24] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Medical Image Computation and Computer Assisted Interventions (MICCAI)*, *Lecture Notes in Computer Science* **9351** (2015) 234–241. ⇒318

[25] J. G. Sled, A. P. Zijdenbos, A. C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, *IEEE Transactions on Medical Imaging* **17**, 1 (1998) 87–97. ⇒320

[26] L. Szilágyi, D. Iclănzan, Z. Kapás, Z. Szabó, Á. Győrfi, L. Lefkovits, Low and high grade glioma segmentation in multispectral brain MRI data. *Acta Universitatis Sapientiae, Informatica* **10** (2018) 110–132. ⇒319, 320, 321, 322

[27] L. Szilágyi, L. Medvés, S. M. Szilágyi, A modified Markov clustering approach to unsupervised classification of protein sequences, *Neurocomputing* **73**, 13–15 (2010) 2332-2345. ⇒325

[28] L. Szilágyi, S. M. Szilágyi, A modified two-stage Markov clustering algorithm for large and sparse networks, *Computer Methods and Programs in Biomedicine* **135** (2016) 15–26. ⇒325

[29] L. Szilágyi, S.M. Szilágyi, B. Benyó, Efficient inhomogeneity compensation using fuzzy c-means clustering models. *Computer Methods and Programs in Biomedicine* **108** (2012) 80–89. ⇒320

[30] S. M. Szilágyi, L. Szilágyi, A fast hierarchical clustering algorithm for large-scale protein sequence data sets, *Computers in Biology and Medicine* **48** (2014) 94–101. ⇒325

[31] A. Tahir, M. Asif, M. Bin Ahmad, T. Mahmood, M.A. Khan, M. Ali, Brain tumor detection using decision-based fusion empowered with fuzzy logic, *Mathematical Problems in Engineering* **2022** (2022) 2710285. ⇒317

[32] U. Vovk, F. Pernuš, B. Likar, A review of methods for correction of intensity inhomogeneity in MRI, *IEEE Trans. Med. Imaging* **26** (2007) 405–421. ⇒320

Received: December 31, 2022 • Revised: January 18, 2023