2021 IEEE AFRICON | 978-1-6654-1984-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/AFRICON51333.2021.9570990

Identifying the most suitable histogram normalization technique for machine learning based segmentation of multispectral brain MRI data

Andrea Kőble, Ágnes Győrfi, Szabolcs Csaholczi, Béla Surányi, Lehel Dénes-Fazakas,
Levente Kovács and László Szilágyi

Abstract— **The main drawback of magnetic resonance imaging (MRI) represents the lack of a standard intensity scale. All observed numerical values are relative and can only be interpreted together with their context. Before feeding MRI data volumes to supervised learning segmentation procedures, their histograms need to be registered to each other, or in other words, they need a so-called normalization. The most popular histogram normalization technique used to assist brain MRI segmentation is the algorithm proposed by Nyúl et al in 2000, which aligns the histograms of a batch of MRI volumes without paying attention to possible focal lesions that might distort the histogram. Alternately, some recent works applied histogram normalization based on a simple linear transform, and reported achieving slightly better accuracy with them. This paper proposes to investigate, which is the most suitable method and parameter settings for histogram normalization to be performed before the segmentation of brain MRI images, separately in the cases of absence and presence of focal lesions.**

Index Terms— **magnetic resonance imaging, tumor segmentation, brain tissue segmentation, histogram normalization.**

I. Introduction

Magnetic resonance imaging (MRI) is a very popular technique in current medical diagnosis, due to its relatively high contrast and fine resolution, and despite its drawback consisting in the fact that recorded numeric values do not directly reflect the observed tissues. The correct interpretation of observed images requires the adaptation of pixel intensities to their context, which is achieved via histogram normalization. The comparison of two intensity values from two different MRI records without having the histograms previously normalized, would be like comparing the value of two jewels by their weight and ignoring to check the precious metal they are made of. Figure 1 presents some brain MRI slices from different records before and after histogram normalization, demonstrating the necessity and the effect of this processing step.

Literature contains several attempts to standardize the intensity distributions of MRI records (e.g. [1], [2], [3], [4], [5]), but none of them were designed to specially handle cases with focal lesions, where the relative intensity of some relevant part of the pixels may seriously differ from the normal. Brain tumors may grow to up to 25% of the brain volume, causing strong distortions in pixel intensity distributions. The most popular histogram normalization technique used by current brain tumor segmentation solutions was proposed by Nyúl *et al.* [1]. This algorithm (referred to as Algorithm A2 in this paper) works in batch mode: it adjusts the intensity distributions of all available records according to the averaged position of some milestones defined as certain percentiles of each input distribution. The more milestones are used, the stronger constraints are applied to the output distributions. However, applying very strong constraints to intensity distributions and expecting them to be similar no matter whether they contain focal lesions or not, may influence the segmentation quality.

The most part of current brain tumor segmentation works (e.g. [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]) only mention that they use the histogram normalization technique of Nyúl *et al.* [1], without giving details of the number of milestones or other parameters. There are few exceptions, where the number of landmark points is revealed: Soltaninejad *et al.* [16] proposed using 12 landmarks, while Pinto *et al.* [17] seem to be using the M12 setting of Algorithm A2, see details in Table I. Alternately, Tustison *et al.* [18] noted that in their study, a simple linear transformation based histogram normalization method provided slightly better accuracy than Algorithm A2, but they did not make their linear transform method public. Other works that employed such linear techniques (e.g. [19], [20], [21]) did not compare their histogram normalization to any other method.

This paper proposes to investigate, which of the above mentioned histogram normalization techniques are most suitable to assist machine learning techniques in achieving best segmentation quality, and what settings are optimal in this process. Two brain MRI segmentation problems are considered separately, one without and the other with focal lesions. In a previous paper [22] we have already compared the effect of various histogram normalization techniques in a brain tumor segmentation problem. This paper can be

This work was supported in part by the Sapientia Institute for Research Programs. This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 679681). The work of L. Szilágyi was supported by the Hungarian Academy of Sciences through the János Bolyai Fellowship program.

A. Kőble, Á. Győrfi, Sz. Csaholczi, B. Surányi and L. Szilágyi are with Computational Intelligence Research Group (CIRG), Sapientia University, Calea Sighişoarei 1/C, 540485 Tîrgu Mureş, Romania (phone: +40-265-206-210; fax: +40-265-206-211; e-mail: lalo at ms.sapientia.ro).

Á. Győrfi, Sz. Csaholczi, B. Surányi and L. Dénes-Fazakas are also with Doctoral School of Applied Mathematics and Applied Informatics (AIAMDI), Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary (phone/fax: +36-1-666-5585; e-mail: gyorfi.agnes at uni-obuda.hu, {szabolcscsaholczi55, bela.suranyi, lehelike} at gmail.com).

L. Kovács and L. Szilágyi are with Biomatics Institute, John von Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary (phone/fax: +36-1-666-5585; e-mail: {kovacs, szilagyi.laszlo} at uni-obuda.hu).

978-1-6654-1984-0/21/$31.00 ©2021 IEEE

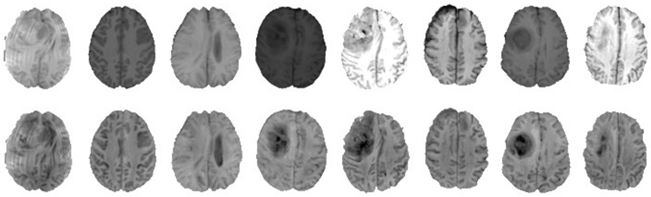

Fig. 1. Eight T1-weighted slices from different records of from the BraTS 2019 LGG data, before (top row) and after histogram normalization (bottom row).

considered an extension of the previous study in the following directions: (1) MRI data with and without focal lesions are considered separately; (2) the arsenal of segmentation techniques involved in the evaluation process is widened; (3) more thorough evaluation is applied; (4) more specific recommendations are formulated in the current study.

II. MATERIALS AND METHODS

A. Data

This study attempts to deal with volumetric MRI image data, both with and without focal lesions. In this order, two data sets are involved in the experiments:

- All ten infant brains from the training data set of the iSeg-2017 challenge [23]. These volumes have two data channels T1 and T2, and the ground truth determined by human experts. The segmentation job requires to distinguish the three main brain tissue types: white matter (WM), grey matter (GM), and cerebro-spinal fluid (CSF).
- Fifty selected low-grade tumor records taken from the Brain Tumor Segmentation (BraTS) challenge training dataset [24], [25], from year 2019. These volumes have four data channels (T1, T2, T1c and FLAIR), and the ground truth determined by human experts. Out of the 76 available records, those 50 were selected for this study, which had missing data only at single pixels. These values were filled with intensity values averaged from intensity values encountered in the $3 \times 3 \times 3$ neighborhood. The considered segmentation job is to separate the whole tumor from normal tissues.

B. Procedure

Both datasets consist of MRI records, so it is not surprising that they are fed to very similar procedures. However, due to the nature of the imaged organs, there are some differences as well. The processing steps are briefly presented in the following:

1) Preprocessing: has the main goal to fill the missing data with averaged values taken from the neighborhood of the pixel in question; histogram normalization, whose techniques are presented in detail in Section III; and feature generation that extends the set of features from 2 to 21 in case of infant brain records, and from 4 to 36 in case of the BraTS records. In both cases, averages from planar neighborhoods of sizes ranging from 3×3 to 11×11, and minimum, maximum and averaged values from spatial $3 \times 3 \times 3$. The feature vector of pixels of the infant brain records are also extended with relative coordinates in the three main directions.

2) Segmentation: via supervised machine learning techniques, which have the goal of distinguishing pixels that belong to different tissue types (WM, GM and CSF in case if infant brain records, lesions and normal tissues in case of BraTS records). Three classification algorithms were included in this study, approaches that functionally strongly differ from each other: random forest (RF), K-nearest neighbors (KNN), and ensembles of support vector machines (SVM). In case of both datasets and all classification approaches, the available data was divided to training and testing data in proportion of 90% vs. 10%. This is explained by the fact that only 10 infant brain records were available that determined us to deploy the "leave-one-out" scheme. We kept the same ratio for the BraTS data as well: the 50 records were randomly divided into ten groups of five, and each group took its turn to serve as testing data while using the other nine groups for training. The classification algorithms were deployed with various settings listed below:

1) RF_1: RF using 2000 randomly chosen pixels from each training volume of infant brains, or 1000 pixels from training volumes of BraTS, allowing maximum tree depth of 18 in both cases.
2) RF_2: RF using 20,000 pixels from each training volume of infant brains, or 10,000 pixels from training volumes of BraTS, maximum tree depth of 22.
3) RF_3: RF using 100,000 pixels from each training volume of infant brains, or 50,000 pixels from training volumes of BraTS, maximum tree depth of 26.
4) KNN_1: KNN using 500 randomly chosen pixels from each training volume of infant brains, or 200 pixels from training volumes of BraTS.
5) KNN_2: KNN using 2000 pixels from each training volume of infant brains, or 1000 pixels from training volumes of BraTS.
6) KNN_3: KNN using 10,000 pixels from each training volume of infant brains, or 5000 pixels from training volumes of BraTS.
7) SVM_1: Ensemble of 15 SVM units, each trained with 30 randomly chosen pixels from each training volume of infant brains or BraTS.

All RF approaches were set to maximum tree count of 45, while all KNN approaches made decisions based on the votes of $k = 11$ neighbors.

3) Postprocessing: is only applied to BraTS records. The outcome of the classification is only considered as intermediary result, which is fed to a morphological post-processing phase. The $11 \times 11 \times 11$ sized neighborhood of each pixel is evaluated: the number of pixels called positive in the intermediate result, and the total number of brain pixels is first extracted from the neighborhood. Those pixels are finally labeled positive, which have a ratio of positives above $1/3$. This post-processing step usually improves the relevant accuracy indicators by 2-3% [20].

4) Statistical evaluation: is employed to find the best performing histogram equalization scheme and settings. In case of infant brain volumes, the main accuracy indicator is the rate of correct decisions ACC, defined as the ratio of pixels whose label match the ground truth. ACC is extracted separately for each volume, and the average of these values is used to characterize the global accuracy of the evaluated methods. In case of whole tumor segmentation, the true positives (TP_i), true negatives (TN_i), false positives (FP_i) and false negatives (FN_i) are first identified for each volume i, $i = 1 \dots 50$. The Dice Score of volume i is defined as $\mathrm{DSC}_i = 2 \times TP_i / (2 \times TP_i + FN_i + FP_i)$. The average Dice Score is then $\mathrm{ADS} = \frac{1}{50} \sum_{i=1}^{50} \mathrm{DSC}_i$, while the overall Dice Score is $\mathrm{ODS} = 2 \times \sum_{i=1}^{50} TP_i / (2 \times \sum_{i=1}^{50} TP_i + \sum_{i=1}^{50} FN_i + \sum_{i=1}^{50} FP_i)$.

III. Methods of Histogram Normalization

Two approaches are compared in this study, both having several applications in current MRI data segmentation methods. The goal is to establish, which approach and what settings are needed to achieve best segmentation accuracy. Both approaches are formulated in such a way that the target set of normalized intensities is the continuous interval $[0, 1]$, which can later be resampled to any desired discrete spectral resolution. Our numerical evaluations use 8-bit coding, using values of 1 to 255 for valid brain pixel intensities and 0 for outer (non-brain) regions of the volume.

A. Method A1: linear transform with one parameter

The linear transform approach works on each volume separately, and treats data channels independently of each other. It transforms any intensity value y to $\min\{\max\{ay + b, 0\}, 1\}$, where coefficients a and b are identified in such a way, that the 25-percentile intensity value p_{25} is transformed to $\lambda_{25} \in (1/4, 1/2)$, while the 75-percentile p_{75} to $1 - \lambda_{25}$. The coefficients of the linear transform are:

$$a = \frac{1 - 2\lambda_{25}}{p_{75} - p_{25}} \quad \text{and} \quad b = \frac{\lambda_{25} p_{75} - (1 - \lambda_{25}) p_{25}}{p_{75} - p_{25}} \ .$$

All input intensity values below $[p_{25}(1 - \lambda_{25}) - p_{75}\lambda_{25}]/(1 - 2\lambda_{25})$ are transformed to 0, while input intensities above $[p_{75}(1 - \lambda_{25}) - p_{25}\lambda_{25}]/(1 - 2\lambda_{25})$ are transformed to 1. So both tails of the input histogram are subject to cutting, but the thresholds are defined dynamically, they depend on the input parameter λ_{25} and the data through percentiles p_{25} and p_{75}. In our previous works [19], [20], [21], this approach was used with parameter setting $\lambda_{25} = 0.4$.

B. Method A2 by Nyúl et al. [1]

The histogram normalization introduced by Nyúl *et al.* [1] treats data channels independently of each other, but uses the chosen data channel of all available volumes to establish the normalized histogram for each volume. It cuts a fixed amount of both tails of input histograms, defined by the percentiles p_{Lo} and $p_{\mathrm{Hi}} = p_{100-\mathrm{Lo}}$, where p_{Lo} is a parameter. Method A2 registers the histograms of different volumes together based on predefined milestones defined as percentiles. Table I presents 12 milestone schemes involved in this study. The step of the algorithm are presented in the following:

TABLE I

Various landmark schemes for the Algorithm A2

Scheme	Landmark points
M01	$p_{\mathrm{Lo}}, p_{50}, p_{\mathrm{Hi}}$
M02	$p_{\mathrm{Lo}}, p_{25}, p_{75}, p_{\mathrm{Hi}}$
M03	$p_{\mathrm{Lo}}, p_{25}, p_{50}, p_{75}, p_{\mathrm{Hi}}$
M04	$p_{\mathrm{Lo}}, p_{10}, p_{50}, p_{90}, p_{\mathrm{Hi}}$
M05	$p_{\mathrm{Lo}}, p_{20}, p_{40}, p_{60}, p_{80}, p_{\mathrm{Hi}}$
M06	$p_{\mathrm{Lo}}, p_{10}, p_{25}, p_{75}, p_{90}, p_{\mathrm{Hi}}$
M07	$p_{\mathrm{Lo}}, p_{20}, p_{35}, p_{50}, p_{65}, p_{80}, p_{\mathrm{Hi}}$
M08	$p_{\mathrm{Lo}}, p_{10}, p_{25}, p_{50}, p_{75}, p_{90}, p_{\mathrm{Hi}}$
M09	$p_{\mathrm{Lo}}, p_{10}, p_{25}, p_{40}, p_{60}, p_{75}, p_{90}, p_{\mathrm{Hi}}$
M10	$p_{\mathrm{Lo}}, p_{10}, p_{25}, p_{40}, p_{50}, p_{60}, p_{75}, p_{90}, p_{\mathrm{Hi}}$
M11	$p_{\mathrm{Lo}}, p_{10}, p_{20}, p_{30}, p_{40}, p_{60}, p_{70}, p_{80}, p_{90}, p_{\mathrm{Hi}}$
M12	$p_{\mathrm{Lo}}, p_{10}, p_{20}, p_{30}, p_{40}, p_{50}, p_{60}, p_{70}, p_{80}, p_{90}, p_{\mathrm{Hi}}$

1) For any record with index h ($h = 1 \dots H$), we establish a bounded linear mapping of original intensities $y^{(h)} \rightarrow \min\{\max\{a^{(h)} y^{(h)} + b^{(h)}, 0\}, 1\}$ in such a way that the percentile p_{Lo} is mapped to 0 and p_{Hi} is mapped to 1. The identified coefficients $a^{(h)}$ and $b^{(h)}$ are used to identify the mapped positions of each milestone $p_m^{(h)}$ ($m = 1 \dots M$, M stands for the number of milestones) from the chosen milestone scheme (see Table I): $p_m^{(h)} \rightarrow \hat{y}_m^{(h)} = a^{(h)} p_m^{(h)} + b^{(h)}$.
2) The final transformation for any record h maps the milestone $p_m^{(h)}$ ($m = 1 \dots M$) to the averaged value $\overline{y}_m = \frac{1}{H} \sum_{i=1}^{H} \hat{y}_m^{(i)}$, and apply piecewise linear interpolation for any intensity value situated between consecutive milestones. So the final transformation of intensity y of record h is given by the formula:

$$y \rightarrow \begin{cases} 0 & \text{if } y < p_{\mathrm{Lo}}^{(h)} \\ 1 & \text{if } y > p_{\mathrm{Hi}}^{(h)} \\ \overline{y}_m + \frac{(\overline{y}_{m+1} - \overline{y}_m)(y - p_m^h)}{p_{m+1}^h - p_m^h} & \text{otherwise} \end{cases} ,$$

where m is established in such a way that $p_m^h \leq y < p_{m+1}^h$.

IV. Results and Discussion

Both histogram normalization methods underwent a thorough evaluation process that involved the two MRI data sets presented in Section II-A, the three classifier algorithms with the 7 settings listed in Section II-B. The statistical evaluation is based on the global accuracy in case of the infant brain MRI data, and the average and overall Dice similarity score in case of the whole tumor segmentation problem. The obtained results are presented in the following paragraphs.

Figure 2 exhibits the global accuracy obtained by the random forest classifier, when using Algorithm A2 for histogram normalization. The plot on the left side shows that at low amount of training data (setting RF_1), the milestone schemes M01, M03, and M07 performed the best, while parameter setting $p_{\mathrm{Lo}} = 0.5\%$ is the optimal choice in most cases. Using larger amount of training data (setting RF_2) leads to higher global accuracy values, while the best performing settings remaining the same. The plot on the right side shows the segmentation accuracy achieved by the setting RF_3 that

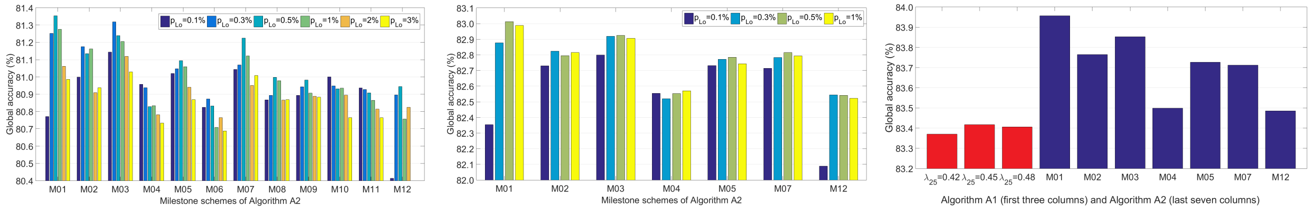

Fig. 2. Global accuracy achieved by the random forest classifiers RF_1 (left), RF_2 (middle), and RF_3 (right) in the segmentation of infant brain tissues, after histogram normalization via Algorithm A2 using various milestone settings and p_{Lo} parameter values, or Algorithm A1 with various values of λ_{25}. RF_3 combined with Algorithm A2 used $p_{\mathrm{Lo}} = 0.5\%$.

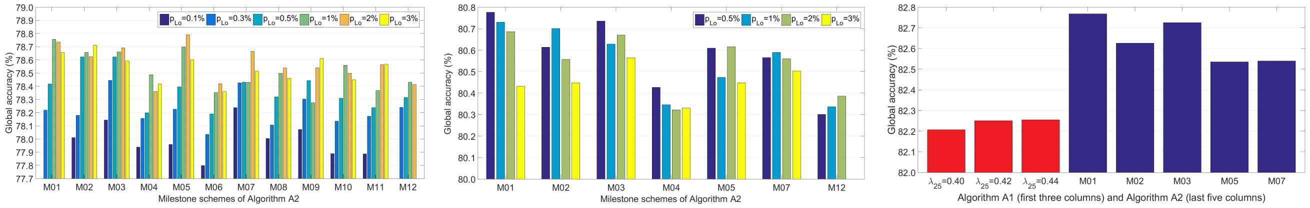

Fig. 3. Global accuracy achieved by the KNN classifiers KNN_1 (left), KNN_2 (middle), and KNN_3 (right) in the segmentation of infant brain tissues, after histogram normalization via Algorithm A2 using various milestone settings and p_{Lo} parameter values, or Algorithm A1 with various values of λ_{25}. KNN_3 combined with Algorithm A2 used $p_{\mathrm{Lo}} = 0.5\%$.

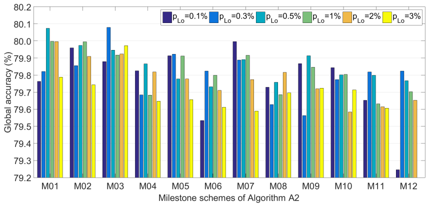

Fig. 4. Global accuracy achieved by the SVM_1 classifier in the segmentation of infant brain tissues, after histogram normalization via Algorithm A2 using various milestone settings and p_{Lo} parameter values.

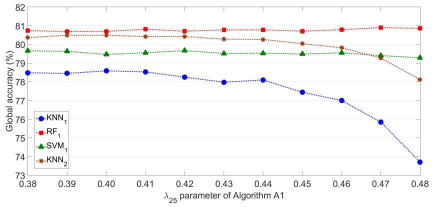

Fig. 5. Global accuracy achieved by the SVM_1, KNN_1, KNN_2 and RF_1 classifiers in the segmentation of infant brain tissues, after histogram normalization via Algorithm A1 using various values of parameter λ_{25}.

uses very large training data set. Here the red columns indicate the accuracy obtained by the use of Algorithm A1 for histogram normalization and the blue ones for Algorithm A2 with $p_{\mathrm{Lo}} = 0.5\%$. Apparently M01 is the best performing milestone scheme, and it gives 0.5% higher global accuracy than Algorithm A1 at any value of λ_{25}. This difference can be seen at lower training data sets as well, if we compare the global accuracy values with the ones presented in Fig. 5.

Analogously to Fig. 2, Fig. 3 presents the segmentation accuracy achieved by the KNN classifier, using three different training data sizes. Here the milestone schemes M01, M02, M03, M05, and M07 seem to perform better than others, and the global accuracy values seem to be 0.2-0.5% higher than in case of linear transform based histogram normalization. Similarly to the RF classifier, KNN seems to perform best when using Algorithm A2 for histogram normalization, M01 or M03 milestone scheme, and $p_{\mathrm{Lo}} = 0.5\%$.

Figure 4 presents the segmentation accuracy results obtained by the ensemble of SVM classifiers using setting SVM_1. The best performing milestone schemes of Algorithm A2 are the same as in case of KNN or RF classifiers. As the best result obtained by SVM in combination with linear transform based histogram normalization is 79.6% (see Fig. 5), Algorithm A2 with its best setting has a lead of 0.5% in accuracy. Figure 5 presents the segmentation accuracy achieved by various classifiers combined with Algorithm A1, plotted against the value of parameter λ_{25}. Whenever $\lambda_{25} < 0.42$, the value of λ_{25} does not influence the accuracy. As λ_{25} approaches its theoretical maximum 0.5, the accuracy achieved by KNN classifier is damaged. This is because at such high values of λ_{25}, hardly anything is cut in the tails of the histograms, and KNN is sensitive to extreme feature values.

For the whole brain tumor segmentation problem, the obtained Dice scores are represented in Figs. 6-9 in an analogous way with Figs. 2-5. The difference here is that instead of a single global accuracy value we have an average and an overall Dice score. The latter is always greater by 2-3%, due to the fact that larger tumors are likely to be detected with better accuracy than smaller ones.

Figure 6 presents the Dice scores obtained by the random forest classifier, when using Algorithm A2 for histogram normalization. The plot on the left side shows that at low amount of training data (setting RF_1), the milestone schemes M01, M02, M03, M05 and M07 performed better than the others, while parameter p_{Lo} should be kept below 0.5% to achieve fine accuracy. Using larger amount of training data (with RF_2 and RF_3) leads to higher global accuracy values,

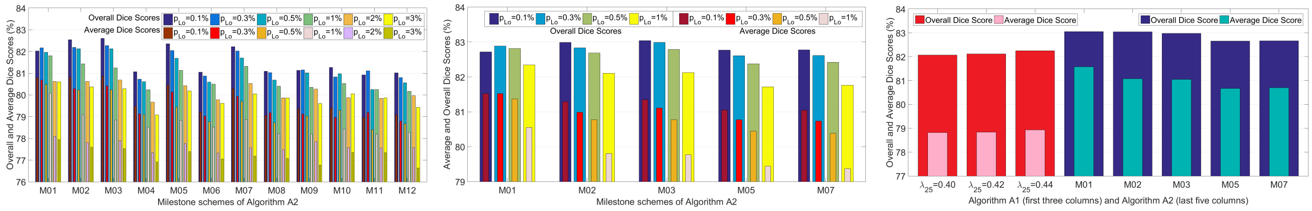

Fig. 6. Average and overall Dice Scores achieved by the random forest classifiers RF_1 (left), RF_2 (middle), and RF_3 (right) in the segmentation of whole tumors, after histogram normalization via Algorithm A2 using various milestone settings and p_{Lo} parameter values, or Algorithm A1 with various values of λ_{25}. RF_3 combined with Algorithm A2 used $p_{\text{Lo}} = 0.5\%$.

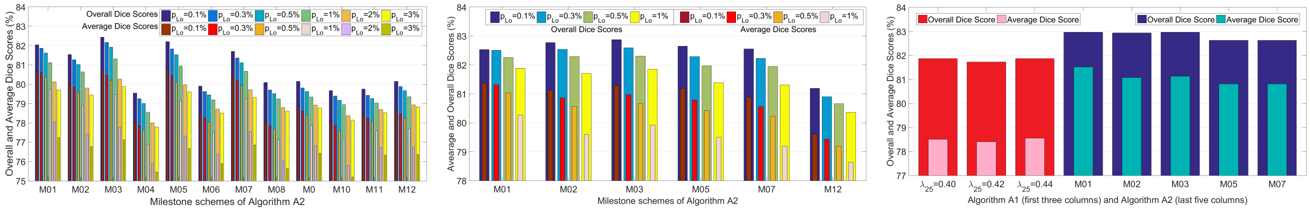

Fig. 7. Average and overall Dice Scores achieved by the random forest classifiers RF_1 (left), RF_2 (middle), and RF_3 (right) in the segmentation of whole tumors, after histogram normalization via Algorithm A2 using various milestone settings and p_{Lo} parameter values, or Algorithm A1 with various values of λ_{25}. KNN_3 combined with Algorithm A2 used $p_{\text{Lo}} = 0.5\%$.

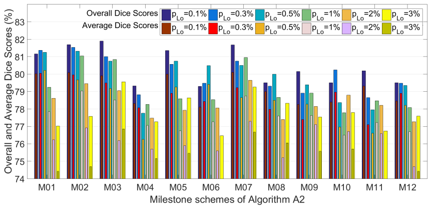

Fig. 8. Global accuracy achieved by the SVM_1 classifier in the segmentation of whole tumors, after histogram normalization via Algorithm A2 using various milestone settings and p_{Lo} parameter values.

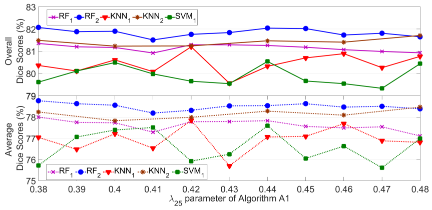

Fig. 9. Average and overall Dice Scores achieved by the SVM_1, KNN_1, KNN_2, RF_1 and RF_2 classifiers in the segmentation of whole tumors, after histogram normalization via Algorithm A1 using various values of parameter λ_{25}.

while the best performing settings prove to be milestone schemes M01, M02 and M03 with p_{Lo} around 0.3%. The plot on the right side also includes three columns that represent best Dice scores obtained with histogram normalization via Algorithm A1, having values 0.8% below the highest ones provided by Algorithm A2.

It is not surprising at all that KNN classifier provides similar results (Fig. 7). Dice scores are somewhat lower than in case of using RF classifier, mainly due to using smaller sets of training data because KNN becomes prohibitively slow when the training data exceeds certain size. KNN gives best Dice scores when combined with Algorithm A2 using milestone scheme M03 and $p_{\text{Lo}} = 0.1\%$.

Figure 8 exhibits the Dice scores obtained by the ensemble of SVM classifiers using setting SVM_1. The best performing milestone schemes of Algorithm A2 are the same ones as in case of KNN or RF classifiers, but not always in the same order. The comparison with the Dice scores achieved using Algorithm A1 for histogram normalization (Fig. 9) reveals that Algorithm A2 at its best settings can provide Dice scores that are greater by 1%. Figure 9 presents the Dice scores achieved by various classifiers combined with Algorithm A1, plotted against the value of parameter λ_{25}. None of the classifiers is really influenced by the value of parameter λ_{25}. The absolute value of the Dice scores are below the ones provided by Algorithm A2.

Recommendations formulated based on the above presented experiments are listed below:

- Algorithm A2, proposed in general form by Nyúl *et al.* [1], can cause better accuracy than a well-designed linear transform in machine learning based segmentation of MRI data, if it is properly adjusted.
- The best performing milestone schemes contain no more than five milestones, including the ones situated at the two ends of the intensity range, p_{Lo} and p_{Hi}.
- The best performing milestone schemes do not use milestones at p_{10} and p_{90} percentiles. The milestones situated closest to p_{Lo} and p_{Hi} should be at least as far as p_{20} and p_{80}, respectively.
- Our experiments showed that p_{Lo} should be set at the 0.5% percentile or below that, and accordingly, p_{Hi} at the 99.5% percentile or above that.
- The above recommendations are valid for histogram normalization of MRI data, no matter whether it contains focal lesions or not.

The results of our investigation suggests that studies like Tustison *et al.* [18] may have achieved up to 1% higher

Dice scores by using the histogram normalization of Nyúl *et al.* [1] with the above recommended settings, instead of deploying the simple linear transform. Similarly, studies like Soltaninejad *et al.* [16] and Pinto *et al.* [17] may have obtained half to one percent higher Dice scores if they used the Algorithm A2 with considerably less milestones. Further on, the uncountable amount of brain MRI segmentation papers within the BraTS mainstream, which only mention having employed the histogram normalization of Nyúl *et al.* [1], may improve their segmentation accuracy by adjusting their work according to our recommendations.

Limitations of this study include the following: only two sets of MRI records were used to investigate the behavior of histogram normalization techniques. Further on, only brain MRI data were involved in this investigation, so its findings may not be equally valid in case of imaging other organs.

V. Conclusions

This paper proposed to investigate, which of the commonly used histogram normalization methods provides the most suitable preprocessed brain MRI data for accurate segmentation via machine learning techniques, and which parameter settings are likely to assure high quality results. Two publicly available brain MRI data sets were involved in the investigation, one without and the other with focal lesions. Based on the results obtained via experimental evaluation, we recommend using the histogram normalization method of Nyúl *et al.* [1], with milestone scheme M03 or M01 given in Table I.

References

[1] L. G. Nyúl, J. K. Udupa, X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.* vol. 19, no. 2, pp. 143–150, 2000.

[2] N. L. Weisenfeld, S. K. Wartfeld, "Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence," *IEEE Int. Symp. on Biomedical Imaging (ISBI)*, Arlington VA, USA, pp. 101–104, 2004.

[3] F. Jäger, Y. Deuerling-Zheng, B. Frericks, F. Wacker, J. Hornegger, "A new method for MRI intensity standardization with application to lesion detection in the brain", in: L. Kobbelt et al (eds), "Vision Model, Visualization", pp. 269–276. AKA GmbH, Köln, Germany, 2006,

[4] K. K. Leung, M. J. Clarkson, J. W. Bartlett, S. Clegg, C. R. Jr. Jack, M. W. Weiner, et al, "Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection," *Neuroimage* vol. 50, pp. 516–523, 2010.

[5] R. T. Shinohara, C. M. Crainiceanu, B. S. Caffo, M. I. Gaitán, D. S. Reich, "Population-wide principal component-based quantification of blood-brain-barrier dynamics in multiple sclerosis," *Neuroimage* vol. 57, pp. 1430–1446, 2011.

[6] S. Pereira, A. Pinto, V. Alves, C. A. Silva, "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI," *1st International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, (BraTS MICCAI 2015, Munich), *Lecture Notes in Computer Science*, vol. 9556, pp. 131–143, 2016.

[7] R. Meier, U. Knecht, T. Loosli, S. Bauer, J. Slotboom, R. Wiest, M. Reyes, M. "Clinical evaluation of a fully-automatic segmentation method for longitudinal brain tumor volumetry," *Scientific Reports*, vol. 6, article no. 23376, 2016.

[8] A. Ellwaa, A. Hussein, E. AlNaggar, M. Zidan, M. Zaki, M. A. Ismail, N. M. Ghanem, "Brain tumor segmantation using random forest trained on iteratively selected patients," *2nd International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (BraTS MICCAI 2016, Athens), *Lecture Notes in Computer Science*, vol. 10154, pp. 129–137, 2017.

[9] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freynmann, K. Farahani, C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, article no. 170117, 2017.

[10] M. Rezaei, K. Harmuth, W. Gierke, T. Kellermeier, M. Fischer, H. J., Yang, C. Meinel, " A conditional adversarial network for semantic segmentation of brain tumor," *3rd International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (BraTS MICCAI 2017, Quebec City), *Lecture Notes in Computer Science*, vol. 10670, pp. 241–252, 2018.

[11] L. Fidon, W. Q. Li, L. C. Garcia-Peraza-Herrera, J. Ekanayake, N. Kitchen, S. Ourselin, T. Vercauteren, "Scalable multimodal convolutional networks for brain tumour segmentation," *Medical Image Computing and Computer Assisted Intervention, vol. III.* (MICCAI 2017, Quebec City), *Lecture Notes in Computer Science*, vol. 10435, pp. 285–293, 2017.

[12] X. Chen, B. P. Nguyen, C. K. Chui, S. H. Ong, "An automated framework for multi-label brain tumor segmentation based on kernel sparse representation," *Acta Polytech. Hung.* vol. 14, no. 1, pp. 25–43, 2017.

[13] W. Chen, B. Q. Liu, S. T. Peng, J. W. Sun, X. Qiao, "Computer-aided grading of gliomas combining automatic segmentation and radiomics," *Int. J. Biomed. Imag.* vol. 2018, article no. 2512037, 2018.

[14] Y. P. Wu, B. Liu, Y. S. Lin, C. Yang, M. Y. Wang, "Grading glioma by radiomics with feature selection based on mutual information," *J. Amb. Intell. Human. Comput.* vol. 9, no. 5, pp. 1671–1682, 2018.

[15] J. Chang, L. M. Zhang, N. J. Gu, X. C. Zhang, M. Q. Ye, R. Z. Yin, Q. Q., Meng, "A mix-pooling CNN architecture with FCRF for brain tumor segmentation," *J. Vis. Commun. Image R.* vol. 58, pp. 316–322, 2019.

[16] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, X. J. Ye, "Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels," *Comput. Meth. Prog. Biomed.*, vol. 157, pp. 69–84, 2018.

[17] A. Pinto, S. Pereira, D. Rasteiro and C. A. Silva, "Hierarchical brain tumour segmentation using extremely randomized trees," *Patt. Recogn.*, vol. 82, pp. 105–117, 2018.

[18] N. J. Tustison, K. L. Shrinidhi, M. Wintermark, C. R. Durst, B. M. Kandel, J. C. Gee, M. C. Grossman and B. B. Avants, "Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR," *Neuroinformics*, vol. 13, pp. 209–225, 2015.

[19] L. Lefkovits, Sz. Lefkovits, and L. Szilágyi, "Brain tumor segmentation with optimized random forest," *2nd International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (BraTS MICCAI 2016, Athens), *Lecture Notes in Computer Science*, vol. 10154, pp. 88–99, 2017.

[20] L. Szilágyi, D. Iclănzan, Z. Kapás, Zs. Szabó, Á. Győrfi, and L. Lefkovits, "Low and high grade glioma segmentation in multispectral brain MRI data", *Acta Universitatis Sapientiae, Informatica*, vol. 10, no. 1, pp. 110–132, 2018.

[21] Á. Győrfi, L. Szilágyi, and 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*, vol. 11, no. 2, art. no, 564, 2021.

[22] Á. Győrfi, Z. Karetka-Mezei, D. Iclănzan, L. Kovács, and L. Szilágyi, "A study on histogram normalization for brain tumour segmentation from multispectral MR image data," *24th Ibero-American Congress on Pattern Recognition*, Havana, Cuba, *Lect. Notes Comp. Sci.* vol. 11896, pp. 375–384, 2019.

[23] L. Wang, D. Nie, G. N. Li, É. Puybareau, J. Dolz, Q. Zhang, *et al.* "Benchmark on automatic 6-month-old infant brain segmentation algorithms: the iSeg-2017 challenge", *IEEE Trans. Med. Imag.* vol. 38, no. 9, pp. 2219–2230, 2019.

[24] 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 Trans. Med. Imag.*, vol. 34, pp. 1993–2024, 2015.

[25] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," arXiv: 1181.02629v3, 23 Apr 2019.