

Article

Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder

Milica M. Badža ^{1,2,*}  and Marko Č. Barjaktarović ² ¹ Innovation Center, School of Electrical Engineering, University of Belgrade, 11120 Belgrade, Serbia² School of Electrical Engineering, University of Belgrade, 11120 Belgrade, Serbia; mbarjaktarovic@etf.bg.ac.rs

* Correspondence: milica.badza@ic.etf.bg.ac.rs; Tel.: +381-11-321-8455

Abstract: The use of machine learning algorithms and modern technologies for automatic segmentation of brain tissue increases in everyday clinical diagnostics. One of the most commonly used machine learning algorithms for image processing is convolutional neural networks. We present a new convolutional neural autoencoder for brain tumor segmentation based on semantic segmentation. The developed architecture is small, and it is tested on the largest online image database. The dataset consists of 3064 T1-weighted contrast-enhanced magnetic resonance images. The proposed architecture's performance is tested using a combination of two different data division methods, and two different evaluation methods, and by training the network with the original and augmented dataset. Using one of these data division methods, the network's generalization ability in medical diagnostics was also tested. The best results were obtained for record-wise data division, training the network with the augmented dataset. The average accuracy classification of pixels is 99.23% and 99.28% for 5-fold cross-validation and one test, respectively, and the average dice coefficient is 71.68% and 72.87%. Considering the achieved performance results, execution speed, and subject generalization ability, the developed network has great potential for being a decision support system in everyday clinical practice.

Keywords: brain tumor segmentation; convolutional neural network; image segmentation; magnetic resonance imaging; machine learning; medical imaging; neural networks



Citation: Badža, M.M.; Barjaktarović, M.Č. Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder. *Appl. Sci.* **2021**, *11*, 4317. <https://doi.org/10.3390/app11094317>

Academic Editors: Il Dong Yun and Soochahn Lee

Received: 24 March 2021

Accepted: 7 May 2021

Published: 10 May 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The advancements in modern technologies, especially in terms of novel machine learning approaches, have an impact on many aspects of everyday life as well as many scientific areas. In the field of medicine, which is a crucial discipline for improving a person's wellbeing, these modern technologies have enabled many possibilities for inspecting and visualizing different parts of the human body. By performing medical screenings, radiologists are able to assess the state of a given body part or organ and base further actions on this knowledge.

In everyday clinical practice, the disease's diagnosis is often based on the analysis of various medical images. Visual observation and interpretation of images can depend on the subjectivity and experience of radiologists. An accurate interpretation of these images can be of great significance since an early diagnosis of a tumor can greatly increase the chances of a successful recovery. Automatic segmentation and classification of multiple changes in images, such as magnetic resonance imaging (MRI) and computer tomography, by computer vision methods, can offer valuable support to physicians as a second opinion [1,2].

Object detection is a widespread problem of computer vision and deals with identifying and localizing a particular class's objects in the image. This aspect of image processing has been an area of interest for researchers for more than a decade. It has found its application in industry, security, autonomous vehicles, as well as in medicine [1].

Computer vision methods for image segmentation can be divided into classical image processing and machine learning approach. Among machine learning approaches, one of the most widely used image processing neural network architectures is convolutional neural networks (CNN). In computer vision, CNN is used for classification and the segmentation of features in the image. Our group already dealt with the issue of classification, and we developed a new CNN architecture for the classification of brain tumors from MRI images from the image database used in this paper [3]. In order to complete the common problems in everyday clinical practice, our group also dealt with the issue of image segmentation for tumor localization, which we described in this paper.

One type of segmentation is semantic segmentation, which aims to classify each pixel of an image into a specific class. The output produces a high-resolution image, usually the same size as the input image. Ronneberger et al. [4] introduced the U-Net for the segmentation of biomedical images in 2015. A U-network is a CNN with autoencoder architecture that restores the probability that an image's pixels belong to certain classes based on the input image and mask, representing an image with accurately classified pixels.

CNN architectures are data-hungry, and in order to train and compare one method with another algorithm, we need a dataset that is adopted by the research community. Several can be found in the literature. For example, the Perelman School of Medicine, University of Pennsylvania, has been organizing a multimodal Brain Tumor Segmentation Challenge (BRATS) [5] online competition since 2012. The image databases used for the competition are small (about 285 images) and are based on two levels of tumors, low and high levels of gliomas, imaged in the axial plane. Image databases are also available after the end of the competition and can be found in papers dealing with the problem of brain tumor segmentation [6–11].

In addition to the database of images from the competition in the works, it is possible to find other different databases available on the Internet [2,12–20] and databases of images collected by the authors [12,21,22]. The largest database of MRI images available on the Internet is also the database used in this paper and contains a total of 3064 images [2]. In comparison, the largest database of images collected by Chang et al. [21] has 5259 images, but it is not available on the Internet. Compared to these two databases, the others contain significantly fewer images, so we choose the dataset that contains 3064 images, since more data reduce the possibility of overfitting.

Tumor segmentation on MRI images is mainly performed by grouping the most similar image pixels into several classes [2,10,16], as well as based on establishing a pixel intensity threshold [23], using super-pixel level features and kernel dictionary learning method [11], and using different CNN architectures [6–9,23,24]. In their paper, Naz et al. [23] presented the architecture of autoencoders for semantic segmentation of brain tumors from images from the same database used in this paper. An autoencoder with an image reduction of four times the input size was realized, and a pixel classification accuracy of 93.61% was achieved. For the same database, Kharrat and Neji [17] achieved a pixel classification accuracy of 95.9% using methods for extracting features from images, selecting the right features using a Genetic Algorithm, and a Simulated Annealing Algorithm. Pixel classification was performed using Support Vector Machines.

In this paper, we present a new convolutional neural autoencoder (CNA) architecture for brain tumor semantic segmentation of three tumor types from T1-weighted contrast-enhanced MRI. The network performance was tested using combinations of two databases (original and augmented), and two different data division methods (subject-wise and record-wise), and two evaluation methods (5-fold cross-validation and one test). The results are presented using the histograms and mean and median values for metrics for pixel classification. A comparison with the comparable state-of-the-art methods is also presented. The best results were obtained by training the network on the augmented database with record-wise data division. The experiment shows that the proposed network architecture has obtained better results than the networks found in the literature trained on the same image database.

2. Materials and Methods

2.1. Image Database

The image database used in this paper consists of 3064 T1-weighted contrast-enhanced MRI images taken at Nanfang and Tianjin General Hospital in China from 2005 to 2010. It was first published on the Internet in 2015, and the last modified version was posted in 2017 [25]. The database contains images showing three types of tumors: meningioma (708 images), glioma (1426 images), and pituitary tumor (930 images). All images were taken on 233 patients in three planes: sagittal (1025 images), axial (994 images), and coronal (1045 images) planes. Examples of different tumor types and different imaging planes are shown in Figure 1a. Tumors are bordered in red. The number of images per patient varies. In addition to MRI images, the database also contains tumor masks, a binary image where 1 indicates a tumor, and 0 everything else, Figure 1b.

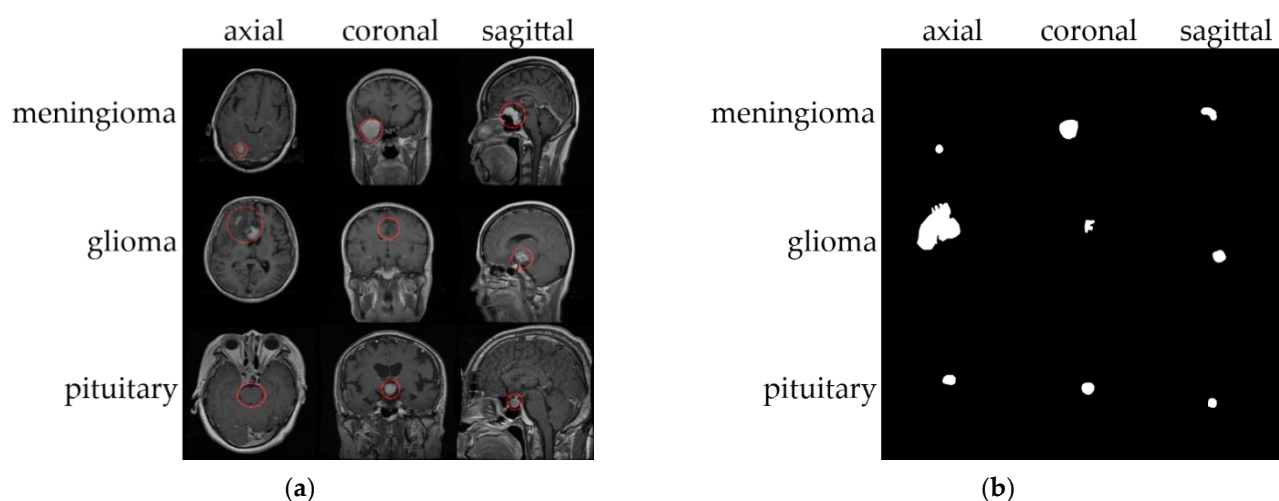


Figure 1. Image example for each type of tumor in each of the planes from the database [25]: (a) Normalized MRI images with bordered tumors; (b) Corresponding tumor masks.

2.2. Image Preprocessing and Data Augmentation

The MRI images found in the database are of different dimensions and are in int16 format. These images represent the network's input layer, so they are normalized to 1 and scaled to 256×256 pixels.

In order to augment the datasets, three separate methods of image transformation were used. The first transformation is a 90° image rotation in a counterclockwise direction, and the second is the flipping of the image about the vertical axis. The last modification of the image is adding the impulse type of noise, the salt and pepper noise. The augmentation is only applied to the training sets. It is augmented four times and consists of the initial, unmodified images and the three modified datasets obtained through the aforementioned modifications methods.

2.3. Network Performance Testing

In order to test the performance of the segmentation network, k -fold cross-validation was used [26]. The cross-validation method divides the database into k approximately equal subsets, one part of which is used for training and validation, and the other for testing. The network is then trained k times, each time using different subsets for training, validation, and testing. In this way, it is achieved that each data is repeatedly found in each set and the impact of imbalanced data split is minimized. Two different evaluation approaches have been implemented, and both are based on 5-fold cross-validation, with one fold being used for testing, one for validation, and the rest for training. The first approach is based on dividing the data into five equal parts so that the images showing each of the tumor categories are equally represented in all parts. This approach is hereinafter referred to as

record-wise cross-validation. The second approach used in testing network performance is based on dividing the data into five equal parts where images from one subject can be found in only one of the parts. Thus, each of the parts contains images of several subjects, regardless of the tumor category. Hereinafter, this approach is called subject-wise cross-validation. A second approach has been implemented to test the network's generalization ability in medical diagnostics [27]. The ability to generalize the network in clinical practice is the network's ability to accurately diagnose data on subjects about which there is no data in the network training process. Therefore, data on individuals in the training set should not appear in the test set. If this is not the case, complex predictors can recognize the interrelationship between identity and diagnostic status and produce unrealistically high classification accuracy [28].

In order to compare the performance of the new network and other existing methods, the network was tested without the use of k -fold cross-validation, i.e., by training the network only once (one test). One test method was used with record-wise and subject-wise approaches for data dividing. When training the network with one test method, data splitting is the same as with the cross-validation methods, i.e., 20% of the data are used for testing, 20% for validation, and 60% for training.

All the methods mentioned above for testing the network were performed on original and augmented datasets. In total, the network was tested using eight tests, combinations of two evaluation methods (5-fold cross-validation and one test), two data division methods (record-wise and subject-wise), and two training datasets (original and augmented).

2.4. Network Architecture

Brain tumor segmentation was performed using a new network architecture with Convolutional layers, developed in Google Colaboratory, with TensorFlow version 2.4.0. and trained on a Graphical Processing Unit (GPU) Tesla P100. The network architecture is modeled based on autoencoders and the U-network [4] and consists of an input, three main blocks, a classification block, and an output, Figure 2. The first main block, Block A, represents the encoder part of the network. It consists of two Convolutional layers that keep the same size at the output as the input one. The reduction in Block A is made using a MaxPool layer that halves the input size. The second block, Block B, consists of two Convolutional layers that retain the input size and serve for additional processing of characteristics, as in the U-network. Block C, the third main block representing the decoder part of the network, is similar to Block A, except that instead of the MaxPool layer, it contains a Transposed-Convolutional layer that doubles the size at the input, followed by two Convolutional layers as in Block A. The classification block has a Convolutional layer and a SoftMax layer that gives the probability to classify each pixel of the image.

All Convolutional layers, except for the one in the classification block, are followed by activation function Rectified Linear Unit and Batch Normalization. These layers were more precisely defined with GlorotNormal [29] kernel initializer and l2 parameter of 0.0001 for kernel regularizer. The Transposed-Convolutional layer was followed by Batch Normalization and was more precisely defined with the same kernel initializer as the Convolutional layers.

The exit from the network represents the binary image that ideally should have the same value for every pixel as the mask. We refer to the network output as a predicted mask. In the end, the segmentation network's architecture consists of an input layer, three Blocks A, Block B, three Blocks C, Classification Block, and Output, and it contains a total of 39 layers and output mask and 488,770 trainable parameters, Table 1.

In the first stage of development, we focused on the U-network with a slight change in the output. The output size was changed to be the same as the input. This network architecture achieved subpar results. The initial modification was to remove the skip connection between the layers from the encoder and decoder parts. The final model architecture, including the number of layers and their parameters, was determined empirically. The proposed network architecture is different from the standard U-network in several aspects.

Firstly, there are no skip connections, there is a smaller number of convolution layers, i.e., blocks, the depth of convolution layers is smaller in the proposed network, and the output of the convolution layer in the proposed network is the same size as the input, unlike the U-net where the output of each convolution layer is 2^2 times smaller than the input. With all the aforementioned differences, the proposed network architecture is smaller than the U-net.

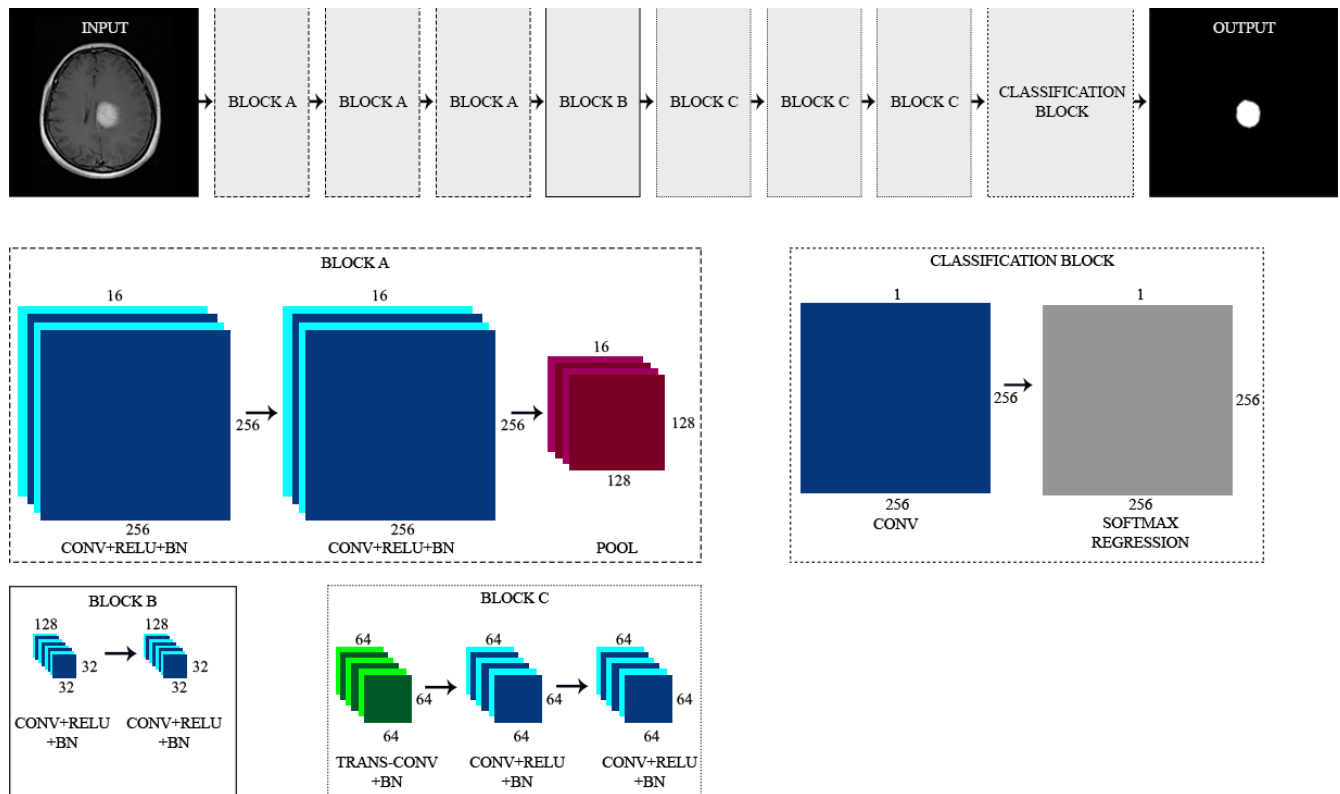


Figure 2. Schematic representation of the CNA segmentation architecture containing the input layer, three Blocks A, Block B, three Blocks C, Classification Block, and Output Layer. Block A represents the encoder, Block C the decoder. Conv layers retain the input dimensions at the output, and Pool and Trans-Conv reduce, i.e., increase the output size. CONV—convolutional layer; RELU—rectified linear unit activation function; BN—Batch Normalization, POOL—association layer; TRANS-CONV—transposed convolutional layer; SOFTMAX—softmax activation layer.

Table 1. New CNA architecture for brain tumor segmentation. All network layers are listed along with their parameters.

Layer No	Layer Name	Block	Layer Properties
1	Image Input	/	$256 \times 256 \times 1$, images
2	Convolutional	Block A	$16 \times 3 \times 3 \times 1$ convolutions with stride [1 1] and padding 'same'
3	Normalization	Block A	Batch Normalization
4	Convolutional	Block A	$16 \times 3 \times 3 \times 16$ convolutions with stride [1 1] and padding 'same'
5	Normalization	Block A	Batch Normalization
6	MaxPool	Block A	2×2 MaxPool with stride [2 2] and padding 'same'
7	Convolutional	Block A	$32 \times 3 \times 3 \times 16$ convolutions with stride [1 1] and padding 'same'
8	Normalization	Block A	Batch Normalization
9	Convolutional	Block A	$32 \times 3 \times 3 \times 32$ convolutions with stride [1 1] and padding 'same'
10	Normalization	Block A	Batch Normalization
11	MaxPool	Block A	2×2 MaxPool with stride [2 2] and padding 'same'
12	Convolutional	Block A	$64 \times 3 \times 3 \times 32$ convolutions with stride [1 1] and padding 'same'
13	Normalization	Block A	Batch Normalization
14	Convolutional	Block A	$64 \times 3 \times 3 \times 64$ convolutions with stride [1 1] and padding 'same'
15	Normalization	Block A	Batch Normalization
16	MaxPool	Block A	2×2 MaxPool with stride [2 2] and padding 'same'

Table 1. Cont.

Layer No	Layer Name	Block	Layer Properties
17	Convolutional	Block B	128 $3 \times 3 \times 64$ convolutions with stride [1 1] and padding ‘same’
18	Normalization	Block B	Batch Normalization
19	Convolutional	Block B	128 $3 \times 3 \times 128$ convolutions with stride [1 1] and padding ‘same’
20	Normalization	Block B	Batch Normalization
21	Transposed-Convolutional	Block C	64 $3 \times 3 \times 128$ convolutions with stride [2 2] and cropping ‘same’
22	Normalization	Block C	Batch Normalization
23	Convolutional	Block C	64 $3 \times 3 \times 64$ convolutions with stride [1 1] and padding ‘same’
24	Normalization	Block C	Batch Normalization
25	Convolutional	Block C	64 $3 \times 3 \times 64$ convolutions with stride [1 1] and padding ‘same’
26	Normalization	Block C	Batch Normalization
27	Transposed-Convolutional	Block C	32 $3 \times 3 \times 64$ convolutions with stride [2 2] and cropping ‘same’
28	Normalization	Block C	Batch Normalization
29	Convolutional	Block C	32 $3 \times 3 \times 32$ convolutions with stride [1 1] and padding ‘same’
30	Normalization	Block C	Batch Normalization
31	Convolutional	Block C	32 $3 \times 3 \times 32$ convolutions with stride [1 1] and padding ‘same’
32	Normalization	Block C	Batch Normalization
33	Transposed-Convolutional	Block C	16 $3 \times 3 \times 32$ convolutions with stride [2 2] and cropping ‘same’
34	Normalization	Block C	Batch Normalization
35	Convolutional	Block C	16 $3 \times 3 \times 16$ convolutions with stride [1 1] and padding ‘same’
36	Normalization	Block C	Batch Normalization
37	Convolutional	Block C	16 $3 \times 3 \times 16$ convolutions with stride [1 1] and padding ‘same’
38	Normalization	Block C	Batch Normalization
39	Convolutional	Classification	2 $3 \times 3 \times 16$ convolutions with stride [1 1] and padding ‘same’
40	Output	/	Two-pixel classes, ‘1’—tumor; ‘0’—other;

2.5. Training Network

The developed architecture for segmentation was trained using Adam optimizers, with a batch size of sixteen images. The initial learning rate was set to 0.0008, and the learning rate scheduler was defined so that every ten epochs learning rate is 0.6 times lower. The number of maximum epochs for network training was set to 300, and the patience for early stopping was 5.

The loss function used for training this network was based on the Sørensen–Dice coefficient (*Dice*), Equation (1),

$$loss = 1 - Dice. \quad (1)$$

The *Dice* coefficient compares the similarities between a segmented image and a mask with marked pixels [30]. The *Dice* coefficient represents the ratio of cross-section and union between two images, Equation (2),

$$Dice = 2 * \frac{|X \cap Y|}{|X| + |Y|}, \quad (2)$$

wherewith $|X|$ is the considerable sum of pixels of the segmented image, and with $|Y|$ the sum of the mask’s pixels.

2.6. Network Performance Metrics

The developed network architecture for tumor segmentation is based on the pixel classification method. Therefore, as the network’s output, we have a binary image (predicted mask), i.e., 256×256 pixels, that could be one of two classes, “1” for tumor and “0” for everything else. In order to evaluate the network performance, we calculated five coefficients

for each image separately. These coefficients are Accuracy (*Acc*), Specificity (*Sp*), Sensitivity (*Se*), Precision (*Pr*) calculated by formulas presented in Equations (3)–(6), respectively,

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (3)$$

$$Sp = \frac{TN}{TN + FP} * 100\% \quad (4)$$

$$Se = \frac{TP}{TP + FN} * 100\% \quad (5)$$

$$Pr = \frac{TP}{TP + FP} * 100\% \quad (6)$$

where *TP* represents true positive pixels, *TN* true negative, *FP* false positive, and *FN* false negative, Table 2.

Table 2. Definition of parameters *TP*, *FP*, *TN*, *FN* between mask pixel values and predicted pixel values.

Parameter	Mask Pixel Value	Predicted Pixel Value
<i>TP</i>	1	1
<i>FP</i>	0	1
<i>TN</i>	0	0
<i>FN</i>	1	0

TP—true positive, *FP*—false positive, *TN*—true negative, *FN*—false negative.

One more coefficient used is the *Dice* coefficient, already mentioned in the previous subsection, Equation (2), only it is transferred to percent.

The calculated coefficients are presented in table format in the Result section, and the histograms could be found in Appendix A. The table shows the mean and median values of each coefficient.

As the segmentation of images is performed by the pixel classification method and as the tumors in the images are insufficient in relation to the remaining images, imbalanced classes exist. Metrics used to keep class imbalances negligible are *Se*, *Pr*, and *Dice* coefficient [31,32].

3. Results and Discussion

The CNA architecture's performance evaluation for segmentation was performed by calculating the *Dice* coefficient and *Acc*, *Se*, *Sp*, and *Pr* for each of the test set images. The coefficients were calculated for each image separately, therefore in Table 3, the mean and median values of those estimated coefficients for all tests are presented. The metrics are calculated and presented in Table 3 for the aforementioned eight tests, combinations of two evaluation methods (5-fold cross-validation and one test), two data division methods (record-wise and subject-wise), and two training datasets (original and augmented).

According to the mean and median value for the *Dice* coefficient, the best results for cross-validation and one test method of the evaluation were obtained by training the network on the augmented dataset with record-wise data division. The observed difference between the mean and the median value is that there are images in which the network could hardly segment the tumor at all, which can be seen on the *Dice* coefficients' histograms in the Appendix A. High values of *Acc* and *Se* coefficients are a consequence of the already mentioned class imbalance. For these reasons, when it comes to image segmentation, it is best to observe the values and histograms of *Dice*, *Sp*, and *Pr* coefficients.

Some of the results of segmentation after training the network on the original data set with record-wise data division are shown in Figure 3. Examples of tumor segmentation with *Dice* coefficients higher and lower from its median value (83.31%) for each tumor type are presented, along with an MRI image and mask. The segmented images also show the

Dice coefficient achieved for that segmentation. Even in the cases where the *Dice* coefficient is lower, the predicted mask clearly indicates the existence and position of the tumor, which shows the significant diagnostic value of our method.

Table 3. Mean and median values of calculated coefficients for network evaluation.

Evaluation Method	Data Division	Training Dataset	Mean/Median	Acc [%]	Se [%]	Sp [%]	Pr [%]	Dice [%]
Cross-validation	Record-wise	Original	mean	99.17	99.68	69.48	70.06	68.78
			median	99.53	99.84	83.89	83.04	82.27
One test	Record-wise	Original	mean	99.22	99.70	70.27	69.02	68.96
			median	99.57	99.85	85.85	83.21	83.31
Cross-validation	Record-wise	Augmented	mean	99.23	99.69	72.31	72.36	71.68
			median	99.58	99.86	86.28	84.51	84.97
One test	Record-wise	Augmented	mean	99.28	99.71	72.65	73.96	72.87
			median	99.60	99.86	86.55	85.93	86.10
Cross-validation	Subject-wise	Original	mean	99.05	99.69	63.75	66.98	63.82
			median	99.49	99.87	79.28	81.45	79.05
One test	Subject-wise	Original	mean	99.21	99.81	65.76	72.15	67.32
			median	99.56	99.89	81.35	85.76	82.45
Cross-validation	Subject-wise	Augmented	mean	99.04	99.62	67.01	65.84	64.74
			median	99.46	99.83	83.69	80.01	79.75
One test	Subject-wise	Augmented	mean	99.17	99.63	71.75	68.70	69.48
			median	99.51	99.81	87.95	84.11	85.09

Acc—accuracy, Se—sensitivity, Sp—specificity, Pr—precision.

The proposed CNA architecture has less than 0.5 million parameters, and training with the augmented dataset achieves better results than training with the original dataset. The record-wise method gives slightly better results compared to the subject-wise. Tumors differ in appearance depending on their type. Additionally, the number of images for each of the tumor types is not the same, nor is the number of patients with each tumor type. Therefore, the subject-wise method does not affect the distribution by type of tumor in the training and testing set, which is the reason for obtaining slightly worse results for the subject-wise method than the record-wise. The segmentation execution time is reasonably good, with an average value of 13 *ms* per image. Despite the advantages the developed network has, there is a drawback worth mentioning. It is regarding the small database currently available and used in this paper. A well known disadvantage of the encoder-decoder architecture is a problem with the slow learning of the middle layer due to gradient decrease with error propagation. The only way to fight this limitation is to make a bigger database.

Comparison with State-of-the-Art-Methods

The proposed CNA architecture's performance was compared with the results presented in papers that used the same database, but different methods and experimental setup. Considering this, the presented comparison is used to put the achieved results into context, since the same database was used. The papers listed in the tables presented their results through the average value of *Acc*, and in order to make a comparison, in Tables 4 and 5 only the coefficient *Acc* is shown. In Table 4 papers using *k*-fold cross-validation for the network testing are presented. Additionally, the performances were compared with works in which the network was not tested by *k*-fold cross-validation but used one test, Table 5. It has been shown that the proposed architecture achieves better results than the results presented in the literature.

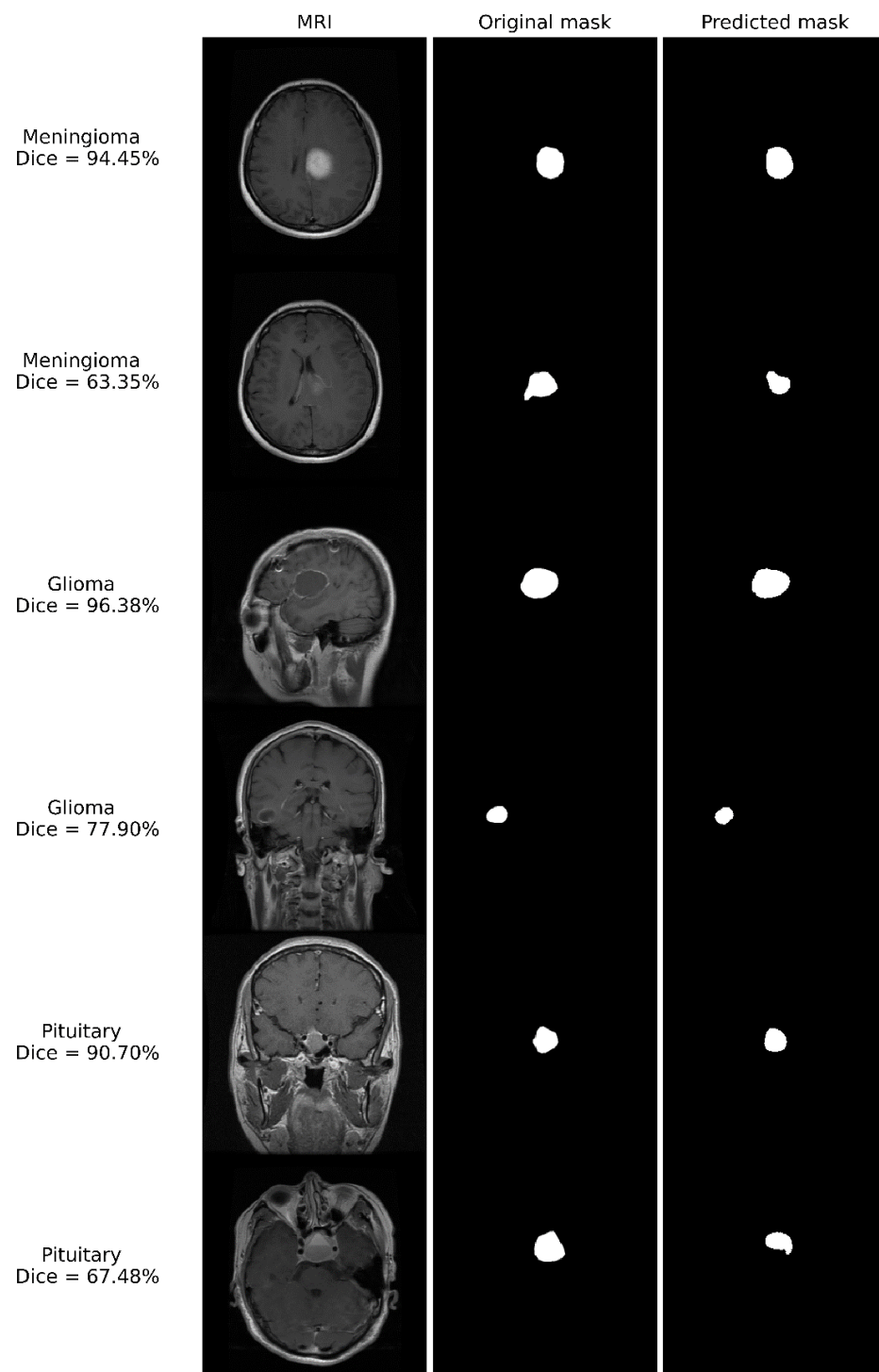


Figure 3. Examples of segmentation after training the network on the original dataset with record-wise data division. The original MRI image, mask, and segmented tumor are shown. The type of tumor, as well as the *Dice* coefficient achieved, are also indicated.

In the literature, we also found other methods that were evaluated on the same dataset, like Chouksey et al. [16]. They have applied a new image segmentation method that involves several different approaches to determine the intensity threshold on which pixels are extracted. Furthermore, they performed segmentation on several other databases, among which was the database used in this paper, but the results show metrics and segmentation for only seven images. Additionally, Kaldera et al. [33] presented the CNN architecture for segmentation and classification of different types of brain tumors, but the

results of segmentation are presented only in the form of a few examples. In his work, Rayhan [1] proposed a new architecture of autoencoders to segment tumors in the brain. Still, when dividing the database, an uneven division was used, i.e., 5% of the images were used for testing, and 0.2% of the training set, which makes up 95% of the total image database, was used the validation set.

Table 4. Comparing the results of different network architectures trained on the original database and performance were tested via *k*-fold cross-validation record-wise method.

Reference	k-Fold Cross-Validation/Data Division	Acc [%]
Kharrat et al. [17]	5-fold; 80% data in training set, 20% in test	95.9
Proposed	5-fold; 60% data in training set, 20% in validation, 20% in test	99.17

Acc—accuracy.

Table 5. Comparing the results of different network architectures trained on the original database and performance were tested via one test record-wise method.

Reference	Data Division	Acc [%]
Naz et al. [23]	70% data in the training set, 15% in validation, 15% in the test	93.61
Proposed	60% data in training set, 20% in validation, 20% in test	99.22

Acc—accuracy.

4. Conclusions

A new CNA architecture for brain tumor segmentation is presented in this study. Tumor segmentation was performed based on semantic segmentation, i.e., classifications of each pixel of the image into two categories, ‘1’ for the tumor, ‘0’ for everything else. The network architecture is simple, with less than 0.5 million parameters. The network was tested using eight tests, combinations of two evaluation methods (5-fold cross-validation and one test), two data division methods (record-wise and subject-wise), and two training datasets (original and augmented).

The best segmentation result for 5-fold cross-validation and one test was achieved with the record-wise method and training the network on the augmented database. The average *Dice* coefficient was 71.68% and 72.87% for 5-fold cross-validation and one test, respectively. There is a difference in results between the increased and the original database. The average *Acc* classification of pixels on the original database is 99.17% and 99.22% for 5-fold cross-validation and one test, respectively, and for augmented it is 99.23% and 99.28%.

To our knowledge, in the literature, no paper tested the generalization of the segmentation network by dividing the data using the subject-wise method for this image database. The best result achieved by dividing the data subject-wise was obtained by training the network on an augmented database. The mean *Acc* classification of pixels by training the network on an augmented image database is 99.04% and 99.17% for 5-fold cross-validation and one test, respectively. The subject-wise data division results are slightly worse than with the record-wise data division, which was expected because the network has no data about the respondent during the testing.

Comparing the new CNA architecture with existing ones that used the same image database and metrically presented the results was also presented. It has been shown that the proposed CNA architecture for segmentation achieves better results than the results presented in the literature. Additionally, the network has good generalization ability, and the execution time required for segmentation is quite good, with an average value of 13 ms per image, implying its effectiveness as an effective decision-support system in order to help medical workers in everyday clinical practice.

In the future, our group plans to combine the already developed CNN architecture for the classification of brain tumors [3] with the presented CNA architecture for segmentation of brain tumors and adapt both networks for use in real-time and real-life conditions

during brain surgery [34], by classifying and accurately localizing the tumor. Since both architectures are small, their adaptation to working in real-time would be possible. It is planned that the developed architecture for segmentation will be tested on other medical image databases and an increased number of subjects. In order to improve the drawbacks, in the future we will expand the dataset with additional images and appropriate segmentation masks. To achieve this, we have started a cooperation with two medical institutes. Additionally, for future work we will consider using other classifiers for the Classification block, such as the classifier based on the KNN algorithm, the enhanced KNN, Support Vector Machine and random forest [35–37].

Author Contributions: Conceptualization, M.M.B.; Methodology, M.M.B. and M.Č.B.; Software, M.M.B.; Validation, M.M.B. and M.Č.B.; Writing—original draft, M.M.B.; Writing—review and editing, M.Č.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank Vladislava Bobić and Ivan Vajs, the researchers at the Innovation Center School of Electrical Engineering, University of Belgrade, and Petar Atanasijević, the teaching assistant at the School of Electrical Engineering, for their valuable comments.

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

Appendix A. Histogram Results

Results after 5-fold cross-validation are also presented using normalized histograms, Figures A1–A8. The histogram, in this case, represents the distribution of the number of images in relation to the range of the obtained calculated metric. In that manner, the histogram preserves the distribution of coefficients by images.

In Figure A1 histogram of the *Dice* coefficient for segmentation results is presented for a test set, after training the network on the original database with 5-fold cross-validation with record-wise division method. The histograms *Acc*, *Se*, *Sp*, and *Pr* for segmentation results on the test set, training network on the original database with 5-fold record-wise cross-validation are presented in Figure A2.

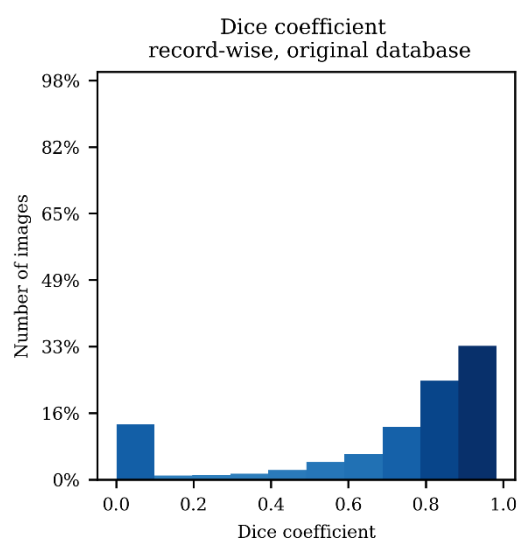


Figure A1. *Dice* coefficient histogram for segmentation results on a test set for a network trained on the original database with 5-fold record-wise cross-validation.

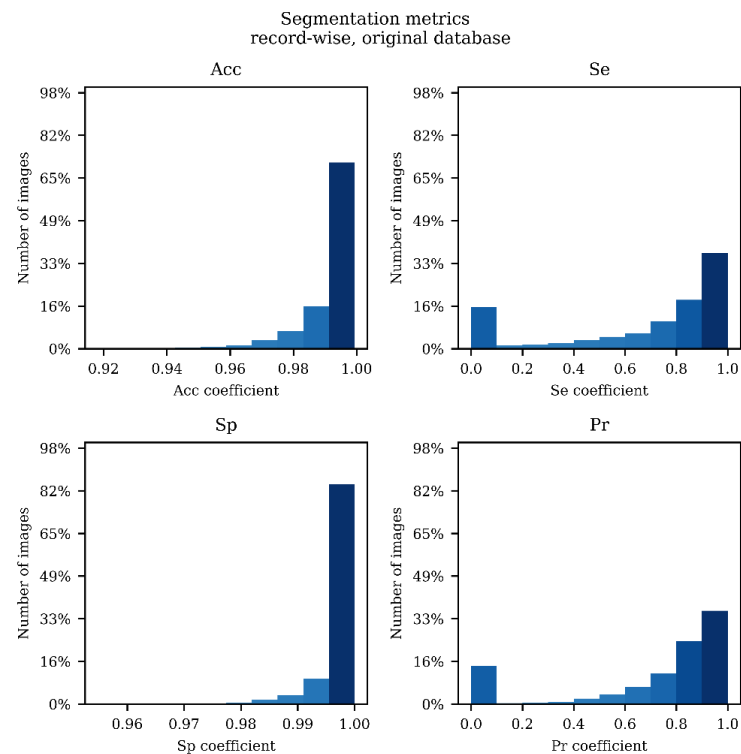


Figure A2. Histograms of metrics for segmentation results on a test set for a network trained on the original database with 5-fold record-wise cross-validation. *Acc*—accuracy; *Se*—sensitivity; *Sp*—specificity; *Pr*—precision.

The histogram of the *Dice* coefficient for the segmentation results on the test set, training the network on the augmented database with 5-fold record-wise cross-validation, is shown in Figure A3.

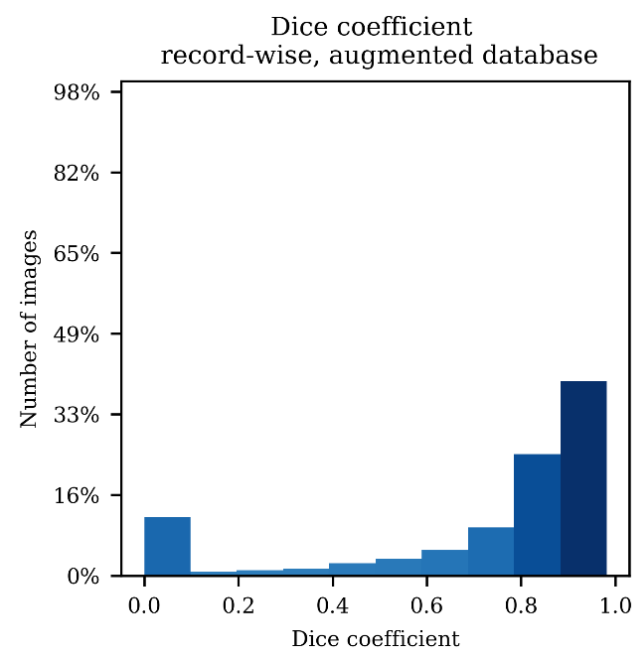


Figure A3. *Dice* coefficient histograms for segmentation results on a test set for a network trained on the augmented database with 5-fold record-wise cross-validation.

In Figure A4, the histograms *Acc*, *Se*, *Sp*, and *Pr* for segmentation results are shown on a test set for a network trained on the augmented dataset with 5-fold record-wise cross-

validation. An improvement of all five coefficients is achieved by expanding the training dataset compared to the original dataset.

Histograms for segmentation results on the test set for the network trained on the original database with 5-fold subject-wise cross-validation are shown in Figures A5 and A6. Observation of the histogram of all five coefficients shows a slight deterioration in relation to the network's training with the record-wise data division.

In Figures A7 and A8, histograms for the *Dice* coefficient and *Acc*, *Se*, *Sp*, and *Pr* for segmentation results are shown on a test set for a network trained on the augmented database with 5-fold subject-wise cross-validation. By increasing the database, improvement of coefficients is achieved compared to the network training on the original database with subject-wise data division. However, it is still slightly worsened than record-wise data division.

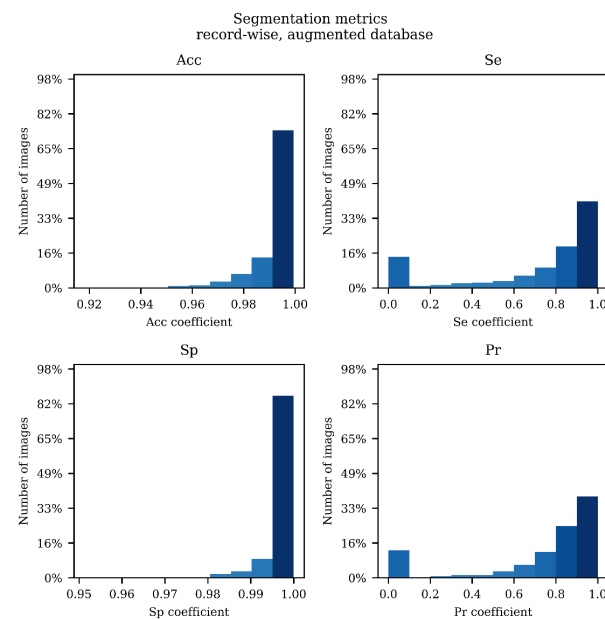


Figure A4. Histograms of metrics for segmentation results on a test set for a network trained on the augmented database with 5-fold record-wise cross-validation. *Acc*—accuracy; *Se*—sensitivity; *Sp*—specificity; *Pr*—precision.

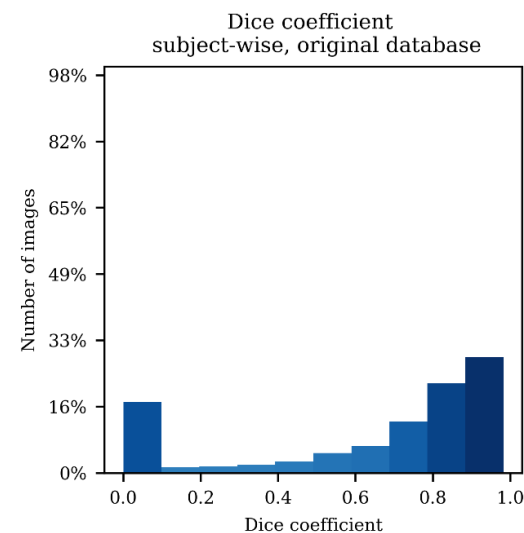


Figure A5. Dice coefficient histograms for segmentation results on a test set for a network trained on the original database with 5-fold subject-wise cross-validation.

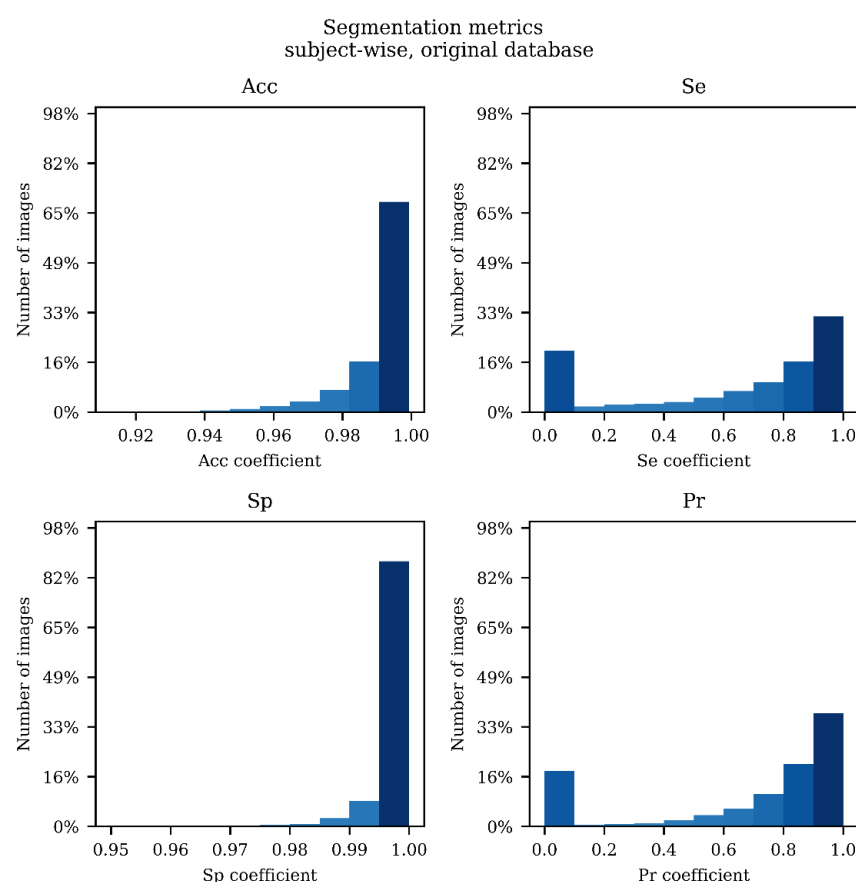


Figure A6. Histograms of metrics for segmentation results on a test set for a network trained on the original database with 5-fold subject-wise cross-validation. *Acc*—accuracy; *Se*—sensitivity; *Sp*—specificity; *Pr*—precision.

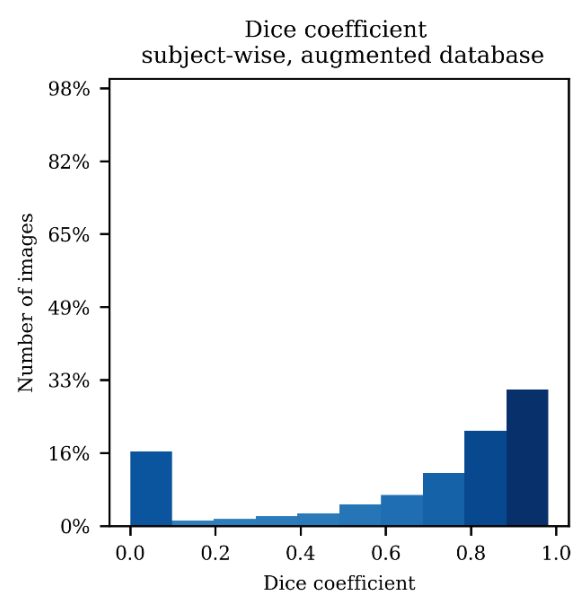


Figure A7. Dice coefficient histograms for segmentation results on a test set for a network trained on the augmented database with 5-fold subject-wise cross-validation.

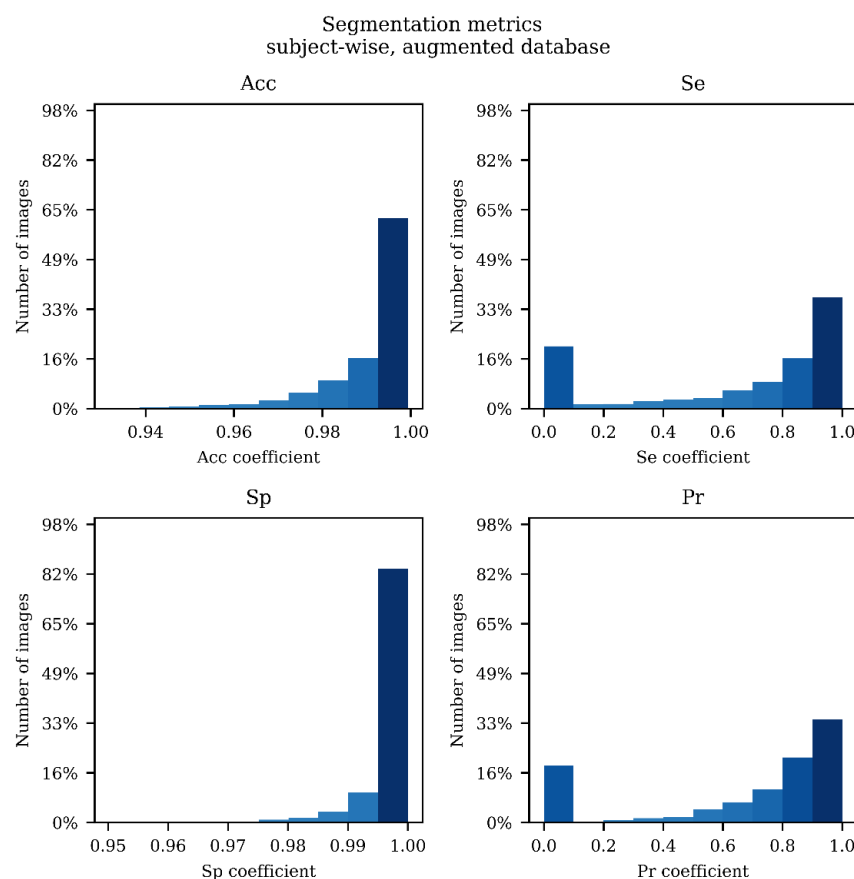


Figure A8. Histograms of metrics for segmentation results on a test set for a network trained on the augmented database with 5-fold subject-wise cross-validation. *Acc*—accuracy; *Se*—sensitivity; *Sp*—specificity; *Pr*—precision.

References

1. Rayhan, F. Fr-Mrnet: A Deep Convolutional Encoder-Decoder for Brain Tumor Segmentation with Relu-RGB and Sliding-Window. *Int. J. Comput. Appl.* **2019**, *975*, 8887.
2. Cheng, J.; Huang, W.; Cao, S.; Yang, R.; Yang, W.; Yun, Z.; Wang, Z.; Feng, Q. Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. *PLoS ONE* **2015**, *10*, e0140381. [\[CrossRef\]](#)
3. Badža, M.M.; Barjaktarović, M.Č. Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. *Appl. Sci.* **2020**, *10*, 1999. [\[CrossRef\]](#)
4. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [\[CrossRef\]](#)
5. Multimodal Brain Tumor Segmentation Challenge BRATS. Available online: <http://braintumorsegmentation.org/> (accessed on 11 March 2021).
6. Jiang, Z.; Ding, C.; Liu, M.; Tao, D. Two-Stage Cascaded u-Net: 1st Place Solution to Brats Challenge 2019 Segmentation Task. In *International MICCAI Brainlesion Workshop*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 231–241. [\[CrossRef\]](#)
7. Saouli, R.; Akil, M.; Kachouri, R. Fully Automatic Brain Tumor Segmentation Using End-to-End Incremental Deep Neural Networks in MRI Images. *Comput. Methods Programs Biomed.* **2018**, *166*, 39–49. [\[CrossRef\]](#)
8. Mlynarski, P.; Delingette, H.; Criminisi, A.; Ayache, N. Deep Learning with Mixed Supervision for Brain Tumor Segmentation. *J. Med. Imaging* **2019**, *6*, 34002. [\[CrossRef\]](#)
9. Amin, J.; Sharif, M.; Yasmin, M.; Fernandes, S.L. Big Data Analysis for Brain Tumor Detection: Deep Convolutional Neural Networks. *Futur. Gener. Comput. Syst.* **2018**, *87*, 290–297. [\[CrossRef\]](#)
10. Amin, J.; Sharif, M.; Raza, M.; Yasmin, M. Detection of Brain Tumor Based on Features Fusion and Machine Learning. *J. Ambient. Intell. Humaniz. Comput.* **2018**, 1–17. [\[CrossRef\]](#)
11. Chen, X.; Nguyen, B.P.; Chui, C.-K.; Ong, S.-H. Automated Brain Tumor Segmentation Using Kernel Dictionary Learning and Superpixel-Level Features. In *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, Hungary, 9–12 October 2016; pp. 2547–2552.

12. Sachdeva, J.; Kumar, V.; Gupta, I.; Khandelwal, N.; Ahuja, C.K. A Package-SFERCB—“Segmentation, Feature Extraction, Reduction and Classification Analysis by Both SVM and ANN for Brain Tumors”. *Appl. Soft Comput.* **2016**, *47*, 151–167. [\[CrossRef\]](#)
13. Javed, U.; Riaz, M.M.; Ghafoor, A.; Cheema, T.A. MRI Brain Classification Using Texture Features, Fuzzy Weighting and Support Vector Machine. *Prog. Electromagn. Res.* **2013**, *53*, 73–88. [\[CrossRef\]](#)
14. Sundararaj, G.K.; Balamurugan, V. An Expert System Based on Texture Features and Decision Tree Classifier for Diagnosis of Tumor in Brain MR Images. In Proceedings of the 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, India, 27–29 November 2014; pp. 1340–1344. [\[CrossRef\]](#)
15. Tripathi, P.C.; Bag, S. Non-Invasively Grading of Brain Tumor Through Noise Robust Textural and Intensity Based Features. In *Advances in Intelligent Systems and Computing*; Springer: Singapore, 2020; pp. 531–539. [\[CrossRef\]](#)
16. Chouksey, M.; Jha, R.K.; Sharma, R. A Fast Technique for Image Segmentation Based on Two Meta-Heuristic Algorithms. *Multimed. Tools Appl.* **2020**, 1–53. [\[CrossRef\]](#)
17. Kharrat, A.; Neji, M. Feature Selection Based on Hybrid Optimization for Magnetic Resonance Imaging Brain Tumor Classification and Segmentation. *Appl. Med Inform.* **2019**, *41*, 9–23.
18. Phaye, S.S.R.; Sikka, A.; Dhall, A.; Bathula, D. Dense and Diverse Capsule Networks: Making the Capsules Learn Better. *arXiv* **2018**, arXiv:1805.04001.
19. Pashaei, A.; Sajedi, H.; Jazayeri, N. Brain Tumor Classification via Convolutional Neural Network and Extreme Learning Machines. In Proceedings of the 2018 8th International Conference on Computer and Knowledge Engineering, ICCKE 2018, Mashhad, Iran, 25–26 October 2018; pp. 314–319. [\[CrossRef\]](#)
20. Sultan, H.H.; Salem, N.M.; Al-Atabany, W. Multi-Classification of Brain Tumor Images Using Deep Neural Network. *IEEE Access* **2019**, *7*, 69215–69225. [\[CrossRef\]](#)
21. Chang, P.; Grinband, J.; Weinberg, B.D.; Bardis, M.; Khy, M.; Cadena, G.; Su, M.-Y.; Cha, S.; Filippi, C.G.; Bota, D. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *Am. J. Neuroradiol.* **2018**, *39*, 1201–1207. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Mohsen, H.; El-dahshan, E.A.; El-horbaty, E.M.; Salem, A.M. ScienceDirect Classification Using Deep Learning Neural Networks for Brain Tumors. *Futur. Comput. Inform. J.* **2018**, *3*, 68–71. [\[CrossRef\]](#)
23. Naz, A.R.S.; Naseem, U.; Razzak, I.; Hameed, I.A. Deep Autoencoder-Decoder Framework for Semantic Segmentation of Brain Tumor. *Aust. J. Intell. Inf. Process. Syst.* **2019**, *15*, 4.
24. Pereira, S.; Meier, R.; Alves, V.; Reyes, M.; Silva, C.A. Automatic Brain Tumor Grading from MRI Data Using Convolutional Neural Networks and Quality Assessment. In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 106–114. [\[CrossRef\]](#)
25. Cheng, J. Brain Tumor Dataset. Available online: https://figshare.com/articles/brain_tumor_dataset/1512427 (accessed on 11 March 2021).
26. Wong, T.T. Performance Evaluation of Classification Algorithms by K-Fold and Leave-One-out Cross Validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [\[CrossRef\]](#)
27. Saeb, S.; Lonini, L.; Jayaraman, A.; Mohr, D.C.; Kording, K.P. The Need to Approximate the Use-Case in Clinical Machine Learning. *GigaScience* **2017**, *6*, 1–9. [\[CrossRef\]](#)
28. Little, M.A.; Varoquaux, G.; Saeb, S.; Lonini, L.; Jayaraman, A.; Mohr, D.C.; Kording, K.P. Using and Understanding Cross-Validation Strategies. Perspectives on Saeb et al. *GigaScience* **2017**, *6*, 1–6. [\[CrossRef\]](#)
29. Glorot, X.; Bengio, Y. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
30. Carass, A.; Roy, S.; Gherman, A.; Reinhold, J.C.; Jesson, A.; Arbel, T.; Maier, O.; Handels, H.; Ghafoorian, M.; Platel, B. Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis. *Sci. Rep.* **2020**, *10*, 1–19. [\[CrossRef\]](#)
31. He, H.; Ma, Y. *Imbalanced Learning*; He, H., Ma, Y., Eds.; Wiley: Hoboken, NJ, USA, 2013. [\[CrossRef\]](#)
32. Haibo, H.; Garcia, E. A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [\[CrossRef\]](#)
33. Kaldera, H.N.T.K.; Gunasekara, S.R.; Dissanayake, M.B. Brain Tumor Classification and Segmentation Using Faster R-CNN. In Proceedings of the 2019 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 26 March–10 April 2019; pp. 1–6. [\[CrossRef\]](#)
34. Moccia, S.; Foti, S.; Routray, A.; Prudente, F.; Perin, A.; Sekula, R.F.; Mattos, L.S.; Balzer, J.R.; Fellows-Mayle, W.; De Momi, E. Toward Improving Safety in Neurosurgery with an Active Handheld Instrument. *Ann. Biomed. Eng.* **2018**, *46*, 1450–1464. [\[CrossRef\]](#)
35. Nguyen, B.P.; Tay, W.-L.; Chui, C.-K. Robust Biometric Recognition from Palm Depth Images for Gloved Hands. *IEEE Trans. Hum. Mach. Syst.* **2015**, *45*, 799–804. [\[CrossRef\]](#)
36. Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [\[CrossRef\]](#)
37. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)