Applied Medical Informatics
Original Research Vol. 42, No. 3 /2020, pp: 180-188

The Use of U-Net Convolutional Neural Network in Magnetic Resonance Images Segmentation

Jan PRZYBYSZEWSKI[*], **Piotr KOHUT**

AGH University of Science and Technology, Department of Robotics and Mechatronics, Mickiewicza 30, Av., 30-059 Krakow, Poland
E-mails: przybysz@student.agh.edu.pl; pko@agh.edu.pl

* Author to whom correspondence should be addressed; Tel.: +48 795027381

Received: July 7, 2020 /Accepted: September 17, 2020/ Published online: September 30, 2020

Abstract
With an aging society, the need to automate time-consuming repetitive actions done by medical doctors to maximize their treatment ability is imminent. Automatic biomedical image segmentation algorithms are set to play a key role in the healthcare of the future. Currently performed by radiologists, the time-consuming procedure consists of assigning areas on the image to corresponding anatomical structures. Novel automatic segmentation algorithms proposed in the literature can be divided into atlas-based, methods using statistical shape knowledge and deep learning algorithms. Deep learning algorithms do not require complex preparation of the atlas or a priori knowledge about the segmented shape. However, their performance is dependent on the training dataset size and quality. Employing the U-Net convolutional neural network architecture, the authors aim to overcome the bottleneck of a small-sized dataset with artificial data augmentation, creating new training samples using flipping and elastic deformation procedures. Algorithms' further increase of efficiency was obtained by combining binary segmentation models – each model was trained to segment one anatomical structure on the image. As most of the work in the field focuses on the introducing novel neural networks' architectures to the field, the thorough description of the impact of these refinement steps sets the paper apart from the other publications in the field. The evaluation of the method utilized Dice's coefficient as a quantitative metric. The presented results show the differences between the model's coefficient values acquired on different magnetic resonance sequences used in the training process. Furthermore, data augmentation impact on segmentation accuracy is showcased, as well as segmentation examples for visual inspection. The authors discuss also the practical usefulness of the algorithm, its limitations as well as future development plans.

Keywords: U-Net; Image Segmentation; Magnetic Resonance Imaging (MRI); Data Augmentation

Introduction

The rapidly increasing quality of biomedical imaging techniques makes them essential in modern healthcare. It is estimated that in a big part of cases, examination of the patient through medical imaging is sufficient for specialists to make an accurate diagnosis. Yet such methods have a drawback: they are processed only by professionals, and the action of reading the images is often time-consuming. Maximizing the efficiency of medical doctors is expected to be a healthcare's key problem in the near future. Europe's aging society indicates that more and more people will need medical treatment. However, with the rising number of potential patients, the number of medical doctors is not increasing. In Poland, in 2019, there were 2.4 medical professionals per 1000 citizens [1]. Taking into account the mentioned society age trend, this number is set to worsen. Current fast development in the sector of artificial intelligence and successful automation of many previously human-covered actions by autonomous algorithms shows promise in solving the above problem. The repetitive tasks of processing biomedical images, currently done by radiologists is an apparent case of activity, which can be optimized by autonomous algorithms. It includes actions such as image classification [2, 3] and image segmentation. For instance, segmentation is currently one of the more time-consuming procedures executed by radiologists. It can be focused on whole anatomical structures [4] or only abnormalities, such as lesions [5]. Image segmentation includes assigning regions of the image to the corresponding structures. The product is a segmentation mask, which contains information about the affiliation of each pixel to the structure. Frequently, the whole sequence of images has to be segmented, which results in hours of processing. The process is often essential in the diagnosis of several diseases (i.e. gonarthrosis) and evaluation of treatment results. Traditional approaches – atlas-based, employ images and reference manual segmentation as an atlas and perform image registration on the processed image [6,7]. Many publications showed robust results using this approach, of which the main advantage is a lower computational resources requirement. Such methods are however, error-prone due to the individual characteristics of patients' anatomy. Recently, more and more research is focused on the deep learning approach to the problem.

Many different, novel convolutional neural networks (CNN) architectures are proposed, not only standard auto-encoder networks, but also dense CNNs, such as the HyperDense-Net [8] or combinations of fully-convolutional network (FCN) and recurrent neural network (RNN) [9]. Large group of researchers combines a deep learning approach with classic algorithms, such as the random walk algorithm, to refine the results [10]. In some cases, the methods are scalable to different modalities [11]. The deep learning approach, contrary to other proposed segmentation methods, such as the methods using statistical shape models [12] or atlas-based, does not need any a priori knowledge about the segmented shape or atlas prepared for reference. It needs, however, a training dataset of images and corresponding segmentation masks, which are prepared manually by specialists. If the number of neural networks' architectures described in the literature is abundant, then there is close to none publications with a strong focus on the impact of dataset augmentation methods and other solutions to the scarcity of the training data. The goal of this study was to develop, train and evaluate a deep learning algorithm, which can be used for the segmentation of human brain magnetic resonance images, with only a sparse dataset. Ideally, the algorithm should segment a sequence of images in a short time period (less than a minute) and obtain a sufficient quality of segmentation.

Material and Method

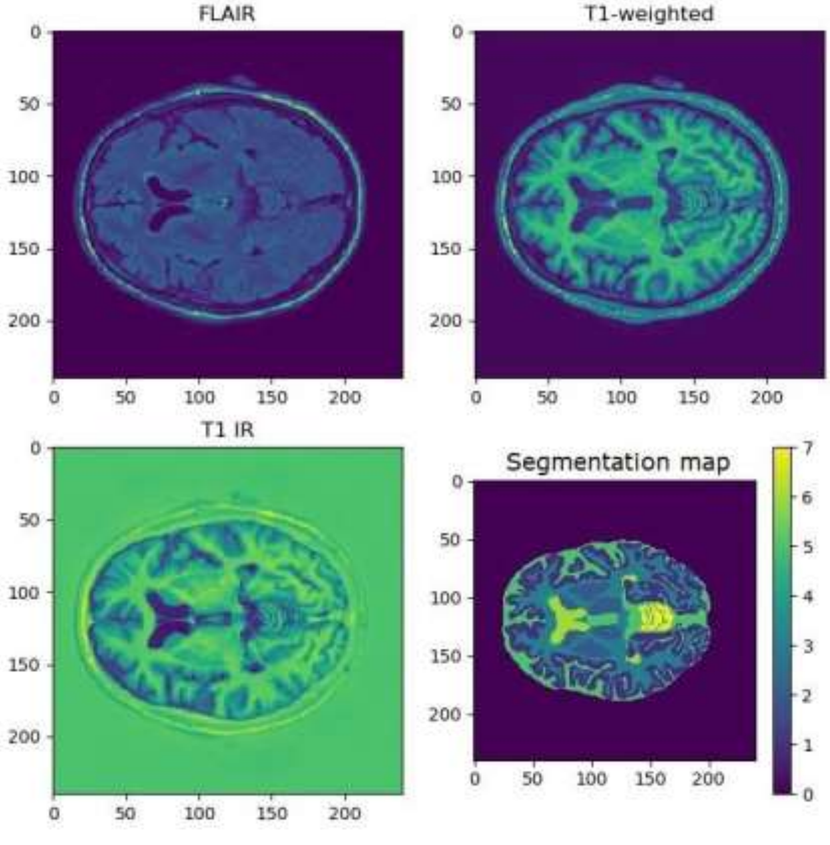
To achieve our goal, the algorithm pipeline contains important preprocessing steps in the form of the dataset augmentation and models binarization. Contrary to the other publications in the field, their impact on the segmentation quality are thoroughly described. These steps in the segmentation algorithms, which use the deep learning approach, can be the key to further improve the quality and create a state-of-the-art algorithm that will be used to streamline segmentation process in the healthcare. With such algorithms complementing other modern solutions, such as the Electronic Health Record [13] and advancements in the e-health sector [14], medical care in the future will have tools to solve the problems that threaten it.

Overall Algorithm Pipeline

The proposed algorithm utilizes human brain magnetic resonance images and a convolutional neural network to segment the image (sequence of images) and return a segmentation mask(s). The convolutional neural network, models' training, evaluation of efficiency and data augmentation were implemented using Python language with Keras, TensorFlow, NumPy, SimpleITK, Elasticdeform and Matplotlib packages.

Dataset

The source of the dataset (Figure 1.) used for training and evaluation was the Grand Challenge on MR Brain Segmentation at MICCAI 2018 [15] with the approval of the competition organizers. It consisted of multi-sequence (T1-weighted, T1-weighted inversion recovery and T2-FLAIR) images obtained on seven patients, with the use of a 3T scanner at the UMC Utrecht. Additionally, every image had a manually segmented mask for reference. Images from different sequences were not mixed up in the process. After visual evaluation, the first and last ten images from each patient's image sequence were excluded, as the shown regions were not valuable for training. That left 196 multi-sequence images in the dataset, which were then split into training and test sets in a 6:1 ratio. Segmentation maps were represented in the form of matrices of the same size as images (240x240x1). Each element of the segmentation map had a value equal to value denoting the anatomical structure, which the corresponding pixel on the image is a part of. Since the available dataset was of small size, it was decided to simplify segmentation by merging labels into 4 structures, instead of 10, present in the original dataset. Merging operations were done as recommended by the dataset creators [16]. The labels convention was as follows: 0 – background, 1 – grey matter, 2 – white matter, 3 – cerebrospinal fluid. Furthermore, data augmentation was applied in two forms. One method created new training samples (image and segmentation mask) by flipping (horizontally, vertically and both) original data. The other method used elastic deformation procedure [17] (Figure 2.). To summarize, the preparation of data resulted in two variants of the training dataset: augmented and not augmented. Each variant consisted of 3 groups of training and test sets, each group using images of a single magnetic resonance sequence. These groups were used to train and test separate models.



FLAIR

T1-weighted

T1 IR

Segmentation map

Figure 1. Original dataset samples, containing images in three MRI sequences – FLAIR (top left), T1-weighted (top right), T1 IR (bottom left) and reference segmentation masks (bottom right). Relation between the color of the structures visible on the mask and structures' labels is denoted by the legend.

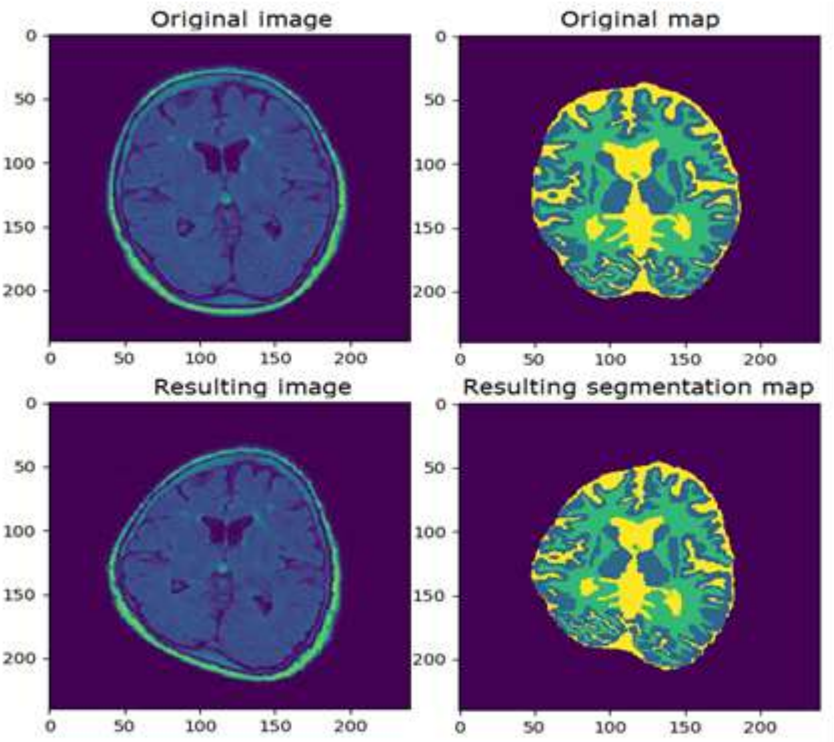


Figure 2. Dataset augmentation using elastic deformation example. Colors in the segmentation maps denote different brain structures. Original image (top left) and its reference segmentation (top right) was distorted by the elastic deformation algorithm, which resulted in a new data sample consisting of a new image (bottom left) and a segmentation map (bottom right).

Net Architecture

The convolutional neural network employed in the algorithm was an adaptation of the U-Net, architecture proposed by Ronneberger et. al. [18], which is capable of fast and precise segmentation, even when trained on compact datasets, which justified its use. The network had two main parts. The first part – encoder, used double convolutional layers interspersed with max-pooling as a downsampling operator. Such structure results in a gradual reduction of width and height and an increase in the depth of the processed image. The following operations were part of the decoder, which utilized double convolution and transposed convolution layers. The selected number of convolutional filters was lower than the depth of the processed volume, as opposed to the number used in the encoder, which combined with upsampling operation resulted in an increase in the height and width of the volume with reduction of depth. Additionally, between double convolution layers in the encoder and the decoder, a skip connection was created [19]. As a result, the input of each double convolution layer in the decoder was a concatenation of the encoder's corresponding layer output and the decoder's previous upsampling layer output. The network was designed to process a 240×240×1 image and return a 240×240×1 segmentation mask, as it was trained to perform binary segmentation.

Models Training

The purpose of the training was to train models specialized in segmenting images of one magnetic resonance sequence. Each model was trained on images of only one sequence. Furthermore, separate models were trained to segment different anatomical structures. Depending on the chosen structure to be segmented by a trained model, segmentation masks from the dataset were first converted to binary form, which used values: 0 – background and other structures, 1 – segmented structure. For instance, one of the models was trained to segment cerebrospinal fluid (CSF) from T1-weighted images, thus the full training was executed using T1-weighted images and segmentation masks converted to a binary form, where 0 denoted pixels that are part of the background, white matter and grey matter. Values equal to 1 denoted CSF. Models were trained using Keras package, with TensorFlow as a backend. Models efficiency was measured using the Dice's coefficient after each epoch on the test dataset (Figure 3). The best model was then saved for evaluation. The tests were conducted on a PC, with i5-7300HQ processor and NVIDIA GTX 1050 graphics card.

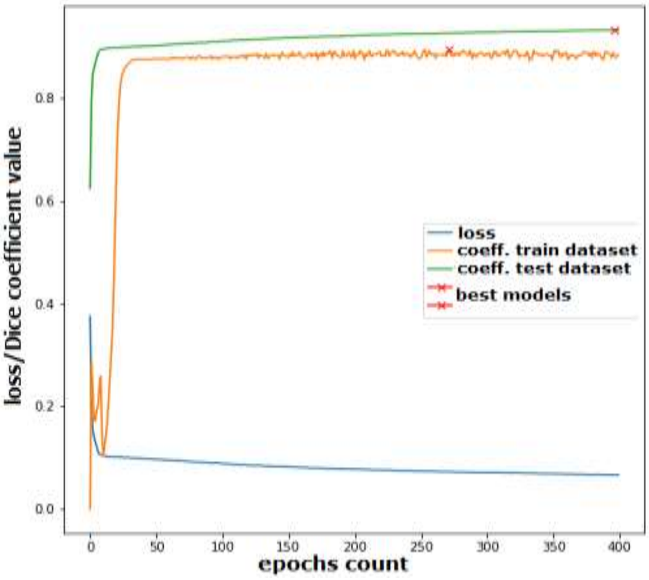


Figure 3. Training process plot

Segmentation Pipeline

The segmentation pipeline of the algorithm utilizes its binary approach by feeding the image to 4 models, each specialized to segment a single structure (Figure 4.). The models return segmentation masks denoting the localization of pixels belonging to each structure. The masks are then merged into a single mask. The value of each pixel is determined using the mask, which assigned it the highest confidence, employing the 4-values label convention. As a result, the returned mask contains the segmentation of all desired structures.

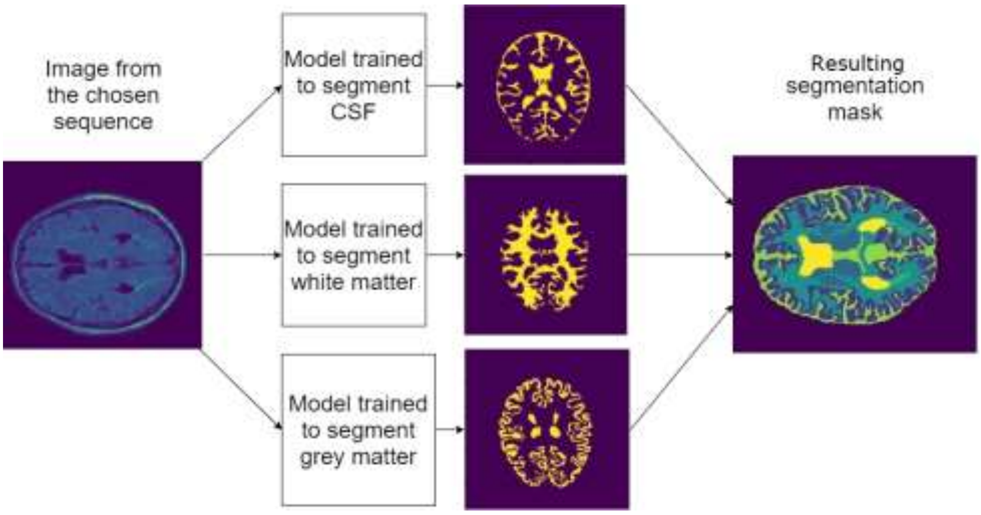


Figure 4. Segmentation algorithm diagram

Statistics

The metric used to evaluate models' performance was Dice's coefficient. During the segmentation process, the X set consists of pixels segmented as a part of the desired structure by the model. Y consists of pixels that are a part of the structure in the reference segmentation mask (ground truth). For binary segmentation, the coefficient is defined as:

$$\frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

where X is the set of pixels denoted as a part of the segmented structure by the model; Y is the true set of pixels being a part of the segmented structure

As for multi-label segmentation masks, the coefficient value was calculated as the mean value of the coefficient for each structure. The coefficient can be interpreted as a similarity measure between two images. If the coefficient value equals 1, then two images are identical. If it equals 0, then there are no common features between the images.

Results

The algorithm took less than half a minute to segment the whole sequence of 28 images. Moreover, utilized training datasets allowed to examine the impact of several factors on the models' effectiveness. There were the following factors: data augmentation methods applied (Figure 5.), type of magnetic resonance sequence (Figure 6.) and segmented anatomical structure. Obtained results show that different magnetic resonance sequences allow models to earn different segmentation effectiveness. Specifically models that were segmenting T1-weighted images, obtained higher Dice's coefficient value on the test dataset, by 8-11% in comparison to models segmenting FLAIR images, and 3-5% in comparison to T1 inversion recovery models. Furthermore, both data augmentation methods allowed for a 3-6% increase in segmentation effectiveness, depending on the structure (Figure 5). Best baseline models (no data augmentation, T1-weighted images) acquired 0.81-0.87 Dice's coefficient, depending on the structure segmented.

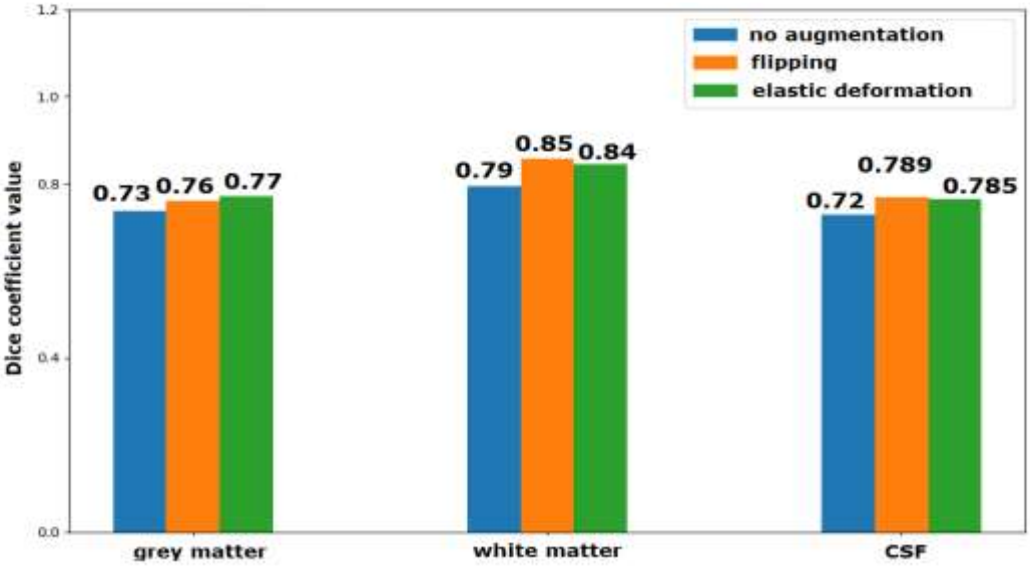


Figure 5. Algorithm's Dice coefficient values with different data augmentation methods used on the training dataset

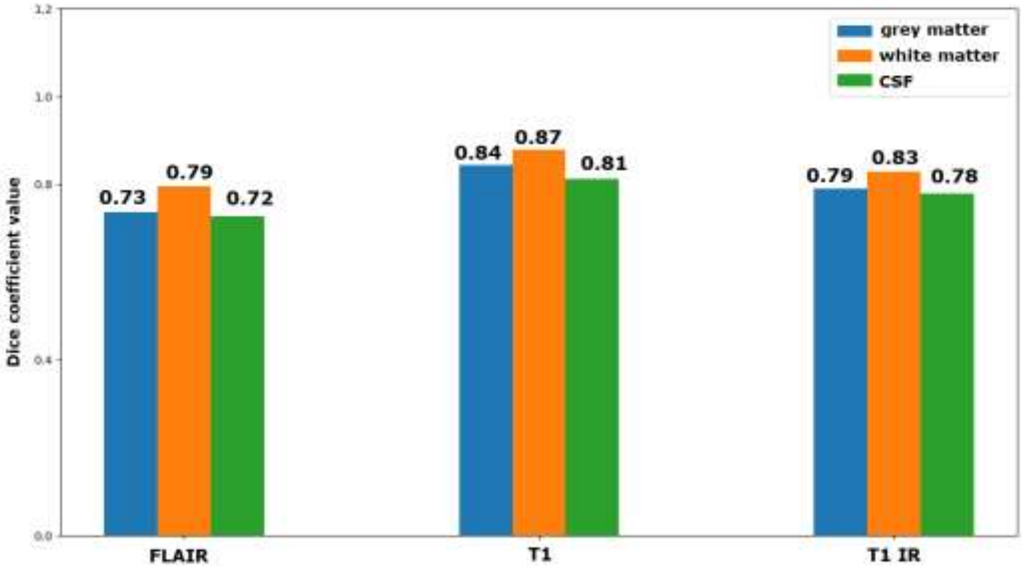


Figure 6. Algorithm Dice coefficient values on different MRI sequences – FLAIR (Fluid Attenuated Inversion Recovery), T1, T1 IR (T1 Inversion Recovery)

Discussion

After model training was completed, the algorithm was capable of segmenting a sequence of images many times faster than a radiology specialist (20-30s). The results reveal that model depends heavily on the MRI sequence used. Models' best performance on T1-weighted images (Figure 6) agrees with specialists' practices, who, according to the dataset creators, were segmenting most structures from T1-weighted images. This indicates that on images of the T1-weighted sequence, the structures are more distinct also for humans. Data augmentation in both presented forms had a beneficial effect on the segmentation efficiency (Figure 5). What is worth noting, simple image flipping technique had the same results as the more sophisticated elastic deformation algorithm. A strong feature of the algorithm is its practical usefulness – it can be run on a common computer setup, with no efficiency issues or longer waiting time. The main limiting factor of the method is the scarcity of the dataset used. To comprehensively test the methods described and take a step towards its implementation in the healthcare, there is a need for another human brain magnetic resonance dataset with reference segmentation masks and ideally taken on different magnetic resonance imaging device, to prove that it is capable of delivering strong results not only on other patients, but also regardless of the hardware used. Other limiting factor may be the U-Net architecture employed, which since its introduction was superseded by novel architectures. There is also an interpretability problem regarding deep learning methods [20]. The planned future studies on the algorithm development include combining it with other segmentation techniques, utilizing statistical shape models, or 3D convolutional neural networks [21,22]. The accuracy may also be increased with the use of feature selection methods [23]. Lately, a big increase in the accuracy has been reported for methods using Conditional Random Field combined with CNNs [24,25]. A rising interest in ensemble models can also be observed [26]. It is important to mention, that currently the biggest bottleneck of all segmentation methods, especially those based on Deep Learning, is the size and the quality of the training dataset. It is very time consuming for the specialists to segment images manually, hence why the datasets available are small and contain at best data from less than 1000 patients. If segmentation algorithms are to be implemented in healthcare, the biggest contributing factor would be to create a larger biomedical image dataset and of better quality. There is however new research conducted in the field of deep learning MRI segmentation, which can create robust models without the prepared expert segmentation [27].

Conclusions

The developed deep learning algorithm was capable of segmentation of human brain biomedical images in a robust manner. The process is quicker than the segmentation conducted by radiologists. Moreover, the impact of data augmentation methods and the MRI sequence used on the score of the algorithm was evaluated. The results show that augmentation methods can have a positive effect on the segmentation quality. However, to thoroughly test the algorithm and fully validate the results, a larger dataset is needed.

References

1. OECD/European Observatory on Health Systems and Policies, Poland: Country Health Profile 2019, State of Health in the EU, OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels. [Internet] 2019 [cited 2020 March 23] Available from: https://doi.org/10.1787/297e4b92-en
2. Mohan M, Sabanayagam S. Detection and Diagnosis of Tumor Regions in Thyroid Images using CANFIS Classifier,.Applied Medical Informatics 2017:39:41-48.
3. Panda A, Tusar KM, Vishnu GP. Automated brain tumor detection using discriminative clustering based MRI segmentation. In: Tiwari S, Trivedi M, Mishra K, Misra A, Kumar K. (eds) Smart Innovations in Communication and Computational Sciences, Smart Innovations in Communication and Computational Sciences. Springer, Singapore. 2019;851:117-126.
4. Aribi Y, Wali A. Kidney Segmentation in Scintigraphic Sequences Data using Multi-Agent Approach. Applied Medical Informatics 2019:2:65-76.
5. Akbarpour T, Shamsi M, Daneshvar S, Pooreisa M. Unsupervised Multimodal Magnetic Resonance Images Segmentation and Multiple Sclerosis Lesions Extraction based on Edge and Texture Features. Applied Medical Informatics 2017:1:30-40.
6. Shan L, Zach C, Charles C, Niethammer M. Automatic Atlas-based Three-label Cartilage Segmentation from MR Knee Images. Proc Workshop Math Methods Biomed Image Analysis 2012, pp. 241-246.
7. Sun L, Zu C, Shao W, Guang J, Zhang D, Liu M. Reliability-based robust multi-atlas label fusion for brain MRI segmentation. Artificial Intelligence in Medicine 2019:96:12-24. Doi: 10.1016/j.artmed.2019.03.004
8. Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ben Ayed I. HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. IEEE Transactions on Medical Imaging 2019:5:1116-1126.
9. Chen J, Yang L, Zhang Y, Alber M, Chen D. Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation. Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain. 2016, pp. 3036-3044.
10. Bao S, Chung A. Multi-scale structured CNN with label consistency for brain MR image segmentation. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 2018:1:113-117.
11. Milletari F, Ahmadi S, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. Computer Vision and Image Understanding 2017:11:92-102.
12. Ambellan F, Tack A, Ehlke M, Zachow S. Automated Segmentation of Knee Bone and Cartilage combining Statistical Shape Knowledge and Convolutional Neural Networks: Data from the Osteoarthritis Initiative. Medical Image Analysis 2019:2:109-118.
13. Moisil I, Vernic C. Old goals revisited. The nursing electronic record. Applied Medical Informatics 2019:41:41.
14. Moisil I, Jitaru E. E-health progresses in Romania. International Journal of Medical Informatics 2006;75(3-4):315-321.
15. MICCAI 2018 Grand Challenge on MR Brain Segmentation [Internet] 2018 [cited 2020 March

23]. Available from: URL: https://mrbrains18.isi.uu.nl/

16. MICCAI 2018 Grand Challenge on MR Brain Segmentation 2018: dataset description [Internet] 2018 [cited 2020 March 23]. Available from: URL:https://mrbrains18.isi.uu.nl/data/
17. Elastic Deform project [Internet] [cited 2020 March 23]. Available from: https://pypi.org/project/elasticdeform/
18. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science 2015:9351:234-241.
19. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The Importance of Skip Connections in Biomedical Image Segmentation. In: Carneiro G. et al. (eds) Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science 2016;10008: 179-187.
20. Marginean A, Groza A, Nicoara SD, Muntean G, Slavescu RR, Letia IA. Towards Balancing the Complexity of Convolutional Neural Network with the Role of Optical Coherence Tomography in Retinal Conditions. 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2019, pp. 475-482, doi: 10.1109/ICCP48234.2019.8959714
21. Singh PS, Wang L, Gupta S, Goli H, Padmanabhan, Gulyás B. 3D Deep Learning on Medical Images: A Review. 2020, doi:arXiv:2004.00218.
22. Kayalibay B, Jensen G, van der Smagt P. CNN-based Segmentation of Medical Imaging Data. 2020, doi: arXiv:1701.03056.
23. Kharrat A, Neji M. Feature selection based on hybrid optimization for magnetic resonance imaging brain tumor classification and segmentation. Applied Medical Informatics 2019:1:9-23.
24. Liu F, Lin G, Shen C. CRF learning with CNN features for image segmentation. Pattern Recognition 2015:10:2983-2992.
25. Kamnitsas K, Ledig C, Newcombe V, Simpson J, Kane A, Menon D, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 2017:2:61-78.
26. Dolz J, Desrosiers C, Wang L, Yuan J, Shen D, Ayed IB. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. Computerized Medical Imaging and Graphics 2020;79:101660. Doi: 10.1016/j.compmedimag.2019.101660
27. Dalca AV, Yu E, Golland P, Fischl B, Sabuncu MR, Iglesias JE. Unsupervised deep learning for Bayesian brain MRI segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 2019, pp. 356-365.