

Automated segmentation of acute stroke lesions using a data-driven anomaly detection on diffusion weighted MRI

Sanaz Nazari-Farsani^{a, *}, Mikko Nyman^b, Tomi Karjalainen^a, Marco Bucci^{a, c}, Janne Isojärvi^a, Lauri Nummenmaa^{a, d}

^a Turku PET Centre and Turku University Hospital, University of Turku, Finland

^b Department of Radiology, Turku University Hospital, University of Turku, Finland

^c Turku PET Centre, Åbo Akademi University, Finland

^d Department of Psychology, University of Turku, Finland

ARTICLE INFO

Keywords:

Acute ischemic stroke
Brain imaging
Automated lesion segmentation
MRI
DWI

ABSTRACT

Background: Successful delineation of lesions in acute ischemic strokes (AIS) is crucial for increasing the likelihood of good clinical outcome for the patient.

New methods: We developed a fully automated method to localize and segment AIS lesions in variable locations for 192 multimodal 3D-magnetic resonance images (MRI) including 106 stroke and 86 healthy cases. The method works based on the Crawford-Howell *t*-test and comparison of stroke images to healthy controls. We then developed a classifier to discriminate the images into stroke or non-stroke categories following the lesion segmentation.

Results: The mean Dice similarity coefficient (DSC) for the test set was 0.50 ± 0.21 (min-max: 0.07 to 0.83) and mean net overlap was 0.66 ± 0.18 (min-max: 0.22–1). The experimental results for the classification of strokes from non-strokes showed mean accuracy, precision, sensitivity, and specificity of 73 %, 0.77 %, 84 %, and 69 %, respectively.

Comparison with existing method: The performance of our methods is comparable with previously published approaches based on machine learning and/or deep learning lesion segmentation techniques. However, most of the previously published methods have yielded low sensitivity, are computationally heavy, and difficult to interpret. The present approach is a significant improvement because it does not require high computation power and memory and can be implemented on a desktop workstation and integrated into the routine clinical diagnostic pipeline.

Conclusions: The current method is straightforward, fast, and shows good agreement with the lesions identified by human experts.

1. Introduction

Magnetic resonance imaging (MRI) and computed tomography (CT) have a critical role in diagnosis, triaging and treatment of acute ischemic stroke (AIS) patients (Powers et al., 2018). CT imaging is the preferred method for patients whose stroke occurred at most 6 h before imaging (Powers et al., 2018). In patients with stroke onset from 6–24 h before imaging, diffusion-weighted imaging (DWI) has been successfully used to select patients for endovascular therapy in recent trials (Albers et al., 2017; Nogueira et al., 2018). The AIS lesions need to be quickly and accurately detected and their volumes must be estimated to triage patients for treatment. Diffusion restriction caused

by acute ischemia can be identified using DWI in minutes after onset and DWI is thus considered the gold standard in ischemic core detection (Vert et al., 2017). This restricted diffusion is visible as elevated DWI and lowered apparent diffusion coefficient (ADC) values (Barber et al., 1998).

Currently, lesion localization is done qualitatively and manually by trained professionals (Fiez et al., 2000; Maier et al., 2015a,b). Manual lesion segmentation may take up to 15 min per case (Martel et al., 2020) and it suffers from inter-operator variability and operator bias (Gillebert et al., 2014; Mah et al., 2014; Neumann et al., 2009). It is thus not optimal in acute stroke management. Automated approaches could, in theory, eliminate inter-subject variability due to their repro-

* Corresponding author at: Turku PET Centre, Kiinamylynkatu 4, Turku, 20520, Finland.
Email address: sanafa@utu.fi (S. Nazari-Farsani)

ducibility and provide fast results for patient care, as well as facilitate automatic analysis of large datasets (Maier et al., 2015a,b).

Recently there has been a growing interest in the automated analysis of neuroradiological images, particularly on AIS lesion segmentation (Maier et al., 2015a,b, Cl'erigues et al., 2019; Perez Malla et al., 2019; Bhanu Prakash et al., 2006; Tsai et al., 2014; Mohd Saad et al., 2011; Menze et al., 2014; Gillebert et al., 2014; Chen et al., 2017; Mah et al., 2014; Boldsen et al., 2018; Stamatakis and Tyler, 2005a; Kamnitsas et al., 2017; Winzeck et al., 2019; Subbanna et al., 2019). The literature on automated and semi-automated stroke lesion segmentation mainly refers to machine learning and neural network methods (Chen et al., 2017; Dou et al., 2016; Mah et al., 2014; Perez Malla et al., 2019; Guerrero et al., 2018; Subbanna et al., 2019; Winzeck et al., 2019; Kamnitsas et al., 2017). Most of these methods have still yielded low sensitivity, are computationally heavy and sometimes difficult to interpret. Currently, widely used commercial software named RAPID uses fixed ADC threshold to segment acute stroke lesions from DWI images with reasonably good results (Straka et al., 2010). Yet, fixed-threshold-method is sensitive to artifacts like T2-blackout, where short T2 values impact even ADC intensity, resulting in false positives for methods relying solely on ADC intensity. Considering all the complications related to the nature of this problem and complexity and inefficiency of the available methods, efficient automatic localization and segmentation of these lesions is still an open challenge.

Here we developed a method, which takes into account both DWI and ADC, to improve lesion segmentation by reducing the number

of false positives caused by fixed ADC thresholding method. The method is based on the Crawford-Howell *t*-test and comparison of stroke images with healthy controls. In comparison with the computationally prohibitive neural network based techniques, our approach imposes low demands for CPU and memory, and can be implemented on an ordinary computer to the routine clinical diagnostic pipeline in order to avoid complications related to manual lesion segmentation. We examined the tendency of the method to make false positives by applying it to non-stroke images (healthy controls). Consequently, we designed a classifier to filter out the real stroke cases from false positive (non-stroke) cases. This filtration step is complementary to segmentation process and plays an important role in reducing the number of false positives.

2. Methods

Our approach was based on three stages: 1) Image preprocessing, 2) Automated lesion segmentation and parameters optimization step, and 3) Image classification based on the segmented lesions (Fig. 1). The accuracy of the automated lesion segmentation was tested against manually delineated lesions drawn by human experts. We tested the tendency of the method to make false positives by applying the automated lesion segmentation method to healthy images. Finally, we trained a classifier to distinguish strokes from non-stroke image masks and evaluated the accuracy of the classifier.

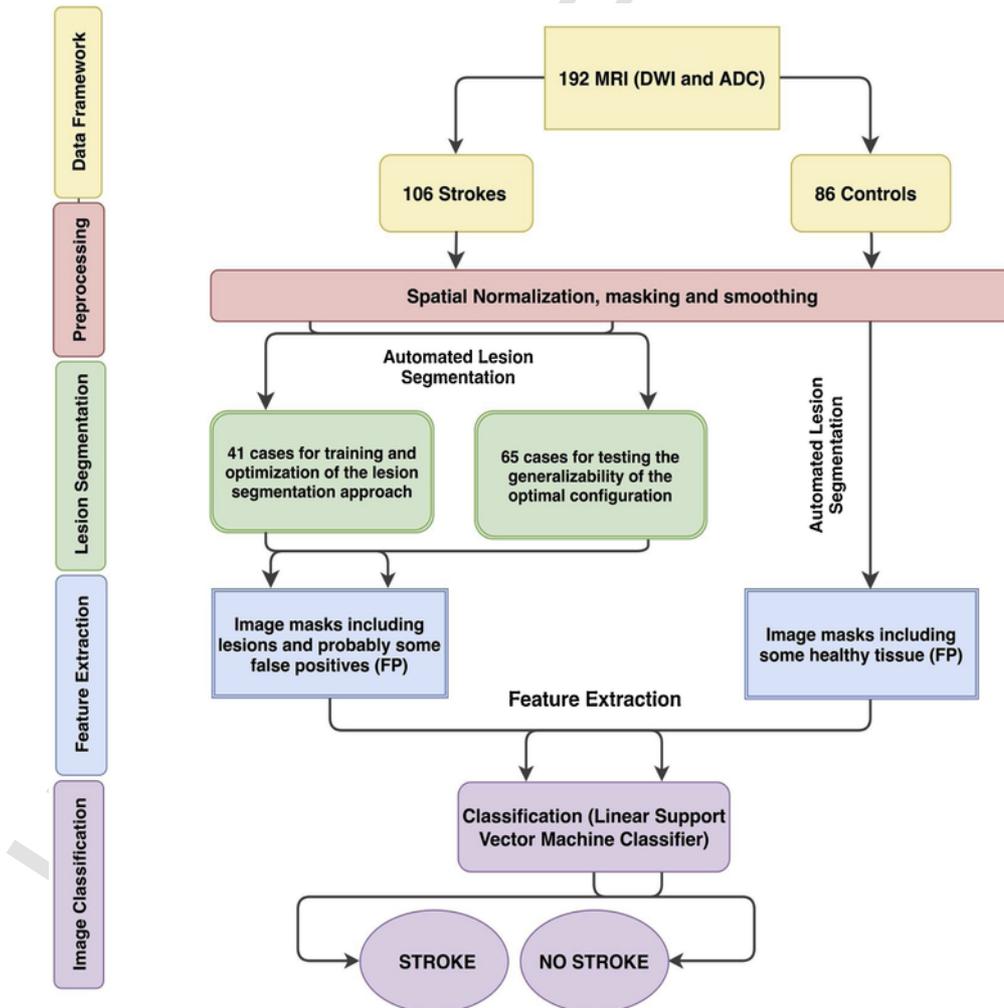


Fig. 1. Overview of the analytic protocol.

2.1. MR image collection and preprocessing

The study protocol was approved by the Hospital District of South-Western Finland. Data (T1 and diffusion-weighted MR images) were analyzed from 192 subjects including 106 patients with AIS and 86 healthy controls. The population included 116 males (mean age of 56.1 years) and 76 females (mean age of 54.5 years). A neuroradiologist carefully screened patients to identify the true AIS cases. All the images were acquired from the Turku University Hospital historical database. For all subjects, the T1 and DW images were acquired on the same imaging session. Images were anonymized prior to analysis. The images were collected from Philips scanners (Philips Ingenia 3T) with the following acquisition parameters: slice thickness: 2.5 or 4 mm, slice spacing: 2.5 or 4 mm, matrix size: 224/128 × 224/128 × (30–50), pixel-size in x–y plane: 1 × 1 or 2 × 2 mm, field strength: 3T, percent phase field of view: 100 or 108, echo time: 85.5–98.78 ms, repetition time: 3112–55091 ms and flip angle: 90°.

Prior to statistical analysis, the images were preprocessed using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) running on MATLAB R2016b. First, the T1 and DW images were aligned using a rigid-object transformation (Collignon et al., 1995). The T1-weighted images were then spatially normalized via the unified segmentation method (Ashburner and Friston., 2005) to a standard anatomical space from the Montreal Neurological Institute (MNI-152). Spatial normalization was crucial in our approach, because the proposed method compares voxels intensities of a given individual to a population-derived distribution of voxels intensities, and individual brains are highly different in terms of size and shape. Accordingly, case and control images were transformed into the common space (MNI) so that they were aligned with one another. The obtained warps were then used for spatially normalizing the DW images. All images were subsequently smoothed using an isotropic Gaussian kernel to account for inaccuracies in normalization. Full width at half maximum (FWHM) of 2–5 mm was tested when optimizing the segmentation parameters. Finally, voxels outside the brain were excluded (masking) as proposed by Abraham and coworkers (Ashburner and Friston, 2005), so that extracranial signal would not confound the analyses.

2.2. Automated lesion segmentation

Our primary endpoint was to develop a method to segment acute infarct lesions automatically. The proposed method works by means of a modified parametric *t*-test called Crawford-Howell, which is designed for the comparison of a single case against a population of controls (Crawford et al., 2009, 2011). In brief, the test assumes that the population distribution of a variable (here MRI signal intensity in a particular voxel in the healthy population), is known and follows a normal distribution. Accordingly, it is possible to calculate how likely each observation is, and improbable events may signal the presence of lesions. The output of this test are *t*-scores calculated as follows:

$$t = \frac{x - \frac{1}{n} \sum_{i=1}^n c_i}{\sqrt{\frac{n+1}{n(n-1)} \sum_{i=1}^n (c_i - \frac{1}{n} \sum_{i=1}^n c_i)^2}} \quad (1)$$

where x is the value for the individual voxel intensity, c_i is the value for the i -th control voxels, and n is the number of control voxels.

Application of this test for the purpose of AIS lesion segmentation has been previously introduced on the CT images (Gillebert et al., 2014). However, the presented method in this study mimics the current golden-standard in DWI-based AIS lesions segmentation by human operators. In this study, the lesion diagnosis is based on the co-occurrence of high-intensity signal on DWI and low-intensity signal on ADC

map. Accordingly, we applied the Crawford-Howell *t*-test separately to the DWI and ADC images. Lesions were identified by thresholding the *t*-scores produced by *t*-tests on DWIs and ADCs in high and low levels, respectively. To reduce the number of healthy tissue voxels which were identified as lesion by the algorithm (false positives), intensity of the voxels selected by *t*-score thresholding were examined once again; if selected voxels by *t*-scores thresholding were among the highest intensity values of the individual's DWI, these voxels were identified as lesion otherwise as healthy tissue. In brief, voxels were identified as lesion if: i) their intensity values were high on DWI and low on ADC compared to healthy population distribution, and ii) the DWI-intensity within identified voxels was high relative to the individual's DWI-intensity distribution. These thresholds (both *t*-scores thresholding and DWI-intensities) were initially assessed empirically on a training set with 41 subjects.

2.2.1. Optimization of the automated method

A total of 3 parameters (thresholds) were optimized for the method; *t*-scores from DWI, *t*-scores from ADC, and intensities of DWI. We defined 8 parameter combinations (2 values for each parameter) of initially discovered thresholds that we thought could result the best. To find out the best parameter combination among these, we divided our data into a training and a test set. We explored different combinations on the training set (41 stroke images) and calculated the spatial overlap between automated and manually delineated lesion masks on this set of data. The combination of parameters producing the best outcome was selected. Next, we examined similarity scores for different smoothing kernels including 2, 3, 4 and 5 mm FWHM to see which FWHM yields the best results. Finally, the performance of the method with optimized parameters and selected FWHM was evaluated on an independent test set (65 stroke cases) to quantify the generalizability of the method.

2.2.2. Lesion delineation by manual operators

The lesions were manually segmented by two researchers who were first trained how to diagnose and segment the lesions by a neuroradiologist (MN). The lesions were manually delineated using Carimas™ 2.9 (Cardiac Image Analysis System) developed at the Turku PET Centre (available for download at www.turkupetcentre.net/carimasturku). DWI and ADC images of each subject were orthonormalized and co-registered in Carimas. Both images were analyzed simultaneously slice-by-slice axially for the lesion detection. The areas appearing bright on DWI and dark on ADC were delineated as lesions. Finally, the manually delineated lesion masks were spatially normalized to MNI space using the transformations from the MR image normalization. Spatial normalization was necessary because the automatic method works with spatially normalized images and for the comparison of manually and automatically created masks. The expert-delineated lesion masks served as ground truth in this study and they were used to calculate the spatial overlap with the automatically created lesion masks. To assess inter-rater and intra-rater agreements, we randomly divided our data to two parts (different from the first round division) and each operator delineated the lesions on her part for the second time.

2.2.3. Validation of the automated lesion segmentation method

Three criteria were used to assess the performance of our lesion segmentation method. First, we calculated Dice similarity coefficient (DSC). The DSC metric measures the spatial overlap between the segmentations; the reference image (A) and estimated image (B) (Zou et al., 2004) and is calculated as:

$$DSC = \frac{2 |A \cap B|}{|A| + |B|} \quad (2)$$

where $|A \cap B|$ is the number of common voxels in images A and B and

$|A|$ and $|B|$ are the number A and B , respectively. DSC has a limited range of $[0,1]$, where DSC of 0 indicates no overlap between the images and DSC of 1 indicates their perfect overlap. Consequently, higher DSC values correspond to the better match of the two segmentations (Zou et al., 2004). DSC is sensitive to false positives. This is particularly problematic when the lesion is small and only contains a few voxels (Chen et al., 2017). Consequently, we also calculated net overlap based on the intersection of automated and manual lesion masks.

Net overlap = (3)

Net overlap essentially measures how well the method can segment the lesions, without penalizing for false positives.

Finally, we calculated the detection rate (DR) of the lesions as described by (Chen et al., 2017) as

$DR = N_{TP}/N$ (4)

where N denotes the number of all the cases and N_{TP} is the number of cases with any true positives (TP) lesion detection.

We also investigated how lesion size and contrast influence these accuracy indices, with the expectation that large and/or high contrast lesions would be more easily detected than small and/or low contrast lesions. We measured the size of the expert-delineated lesion mask (ground truth) volumes and calculated the correlation between volume sizes and similarity scores. To evaluate the effect of contrast, we first created mirror masks from manually drawn lesions and used those as references. Reference masks were monitored and two cases with the expansion of the lesion in both hemisphere were excluded for this part of the analysis. We calculated the average intensity of the both lesion masks (I_{lesion}) and reference masks (I_{ref}). The ratio of the intensities (I_{lesion} / I_{ref}) served as an indicator of the contrast of the individual lesions. Consequently, we calculated the correlation between intensity ratios and similarity scores.

2.3. Image classification

Our secondary aim was to develop a classifier that could distinguish stroke cases from non-stroke ones. To this end, besides the 106 stroke cases, we applied our automated lesion segmentation method to 86 healthy MR images. Ideally, the method should not identify any voxels as an acute lesion in healthy humans, but in practice, the method produces some false positives in non-stroke images. To refine the results of the automated segmentation step, we developed a classifier to categorize stroke cases from non-stroke ones. As input, the classifier uses the image masks created by the automated lesion segmentation method from both stroke cases and healthy controls.

2.3.1. Feature extraction from image masks

We extracted several features from the image masks of both stroke cases and non-stroke and fed them to the classifier. We expected that signal intensity and geometry of the image masks could be used to classify the masks. Therefore, we made two main categories of features; intensity-based features and geometric features. Intensity features included mean, median, standard deviation, skewness, kurtosis and entropy of the DW image mask's intensities. Geometric features were mainly image region properties such as area, centroid and bounding box of the image masks (Legland et al., 2011).

2.3.2. Classifier algorithm, optimization and evaluation

The classifier was used to categorize the cases into either stroke or non-stroke. We set up a support vector machine (SVM) classifier with a linear kernel. To optimize the penalty parameter and to evaluate the performance of the classifier, we applied nested cross-validation

(CV) (Cawley and Talbot, 2010). First, we split our data into four folds (outer loop), run the inner cross-validation (four-folds) on three of the folds (merged together) and evaluated the model on the fourth fold. This was repeated four times so that every fold was used for testing once. In this way, by inner loop, the optimal penalty parameter was assessed and the independent testset from the outer loop was utilized to get an unbiased estimation of the classifier performance. To examine the stability of the method, the procedure was repeated over 30 iterations.

3. Results

Fig. 2 illustrates the segmentation algorithm applied on three different patients. Fig. 2(A) shows the two corresponding inputs: DWI and ADC variables. Fig. 2(B) shows the following preprocessing phase where we applied spatial normalization, co-registration, and smoothing. Fig. 2(C) shows the resulting segmented lesions delineated by the experts (red regions) and propose algorithm (blue regions). In Fig. 2(D), threshold optimization is made by fine-tuning the statistical parameters. Finally, the overall overlap (DSC) between the expert and algorithm delineated segmented regions is calculated (Fig. 2(E)).

The results of the DSC (overall overlap) and net overlap of the automated and manual lesion masks for the 8 different configurations (sets of thresholds) are shown in Table 1. The third configuration yielded the best net overlap (mean = 0.59) and the eighth configuration resulted in best DSC (mean = 0.50) on the train set. Fig. 3 shows several examples of lesion segmentation obtained by the third and eighth configurations. After finding the optimal configurations, we examined the results for the smoothing values ranges from 2 to 5 mm FWHM (Fig. 4(A)) in the training set. In general, smoothed images with FWHM of 3 mm yielded the best results (DSC = 0.50 and net overlap = 0.61). Finally, we tested the generalizability of the selected configurations and FWHM on an unseen test set ($n = 65$). The average DSC on the test set was 0.50 (standard deviation = 0.21) and mean net overlap was 0.66 (standard deviation = 0.18) (Fig. 4(B)).

The median volume of the lesions delineated by two readers for inter-rater correlation analysis was 3.32 cm³, ranging from 0.19–82.59 cm³ (IQR: 15.29) and 5.89 cm³ for the lesions delineated by the same operators two times for the intra-rater correlation analysis, range from 0.10–239.40 cm³ (IQR: 17.98). The inter-rater and intra-rater correlations coefficient were 0.949 ($P < 0.001$) and 0.996 ($P < 0.001$), respectively. However, volume-based method is not the optimal approach for the similarity assessment of the lesion segmentation, because volume discards the location information. For example, the lesion masks may have similar volumes but different shapes. To further qualify the consistency of the lesion delineations, we calculated the DSC between the lesion masks delineated on both rounds. This yielded a mean DSC of 0.77, with standard deviation of 0.12. This analysis further suggests that the manual lesion segmentation suffers from operator-bias and lack of reproducibility.

The detection rate of our method in the test data was 100%; meaning that in all the stroke cases the method identified some true positives (TP) with net overlap ranging from 0.13 to 1.

Fig. 5 shows the linear relationship between the spatial overlap metrics and the lesion size and/or intensity ratios. The Pearson correlation coefficient between spatial overlap metrics and size of the lesion was 0.26 ($P = 0.007$) for the DSC and 0.21 ($P = 0.034$) for the net overlap. The correlation coefficient between spatial overlap and intensity ratio were also 0.42 ($P < 0.001$) and 0.30 ($P = 0.002$) for the DSC and net overlap respectively.

The experimental results for the classification of strokes from non-strokes showed mean accuracy, precision, sensitivity, and specificity of approximately 73%, 77%, 84%, and 69%.

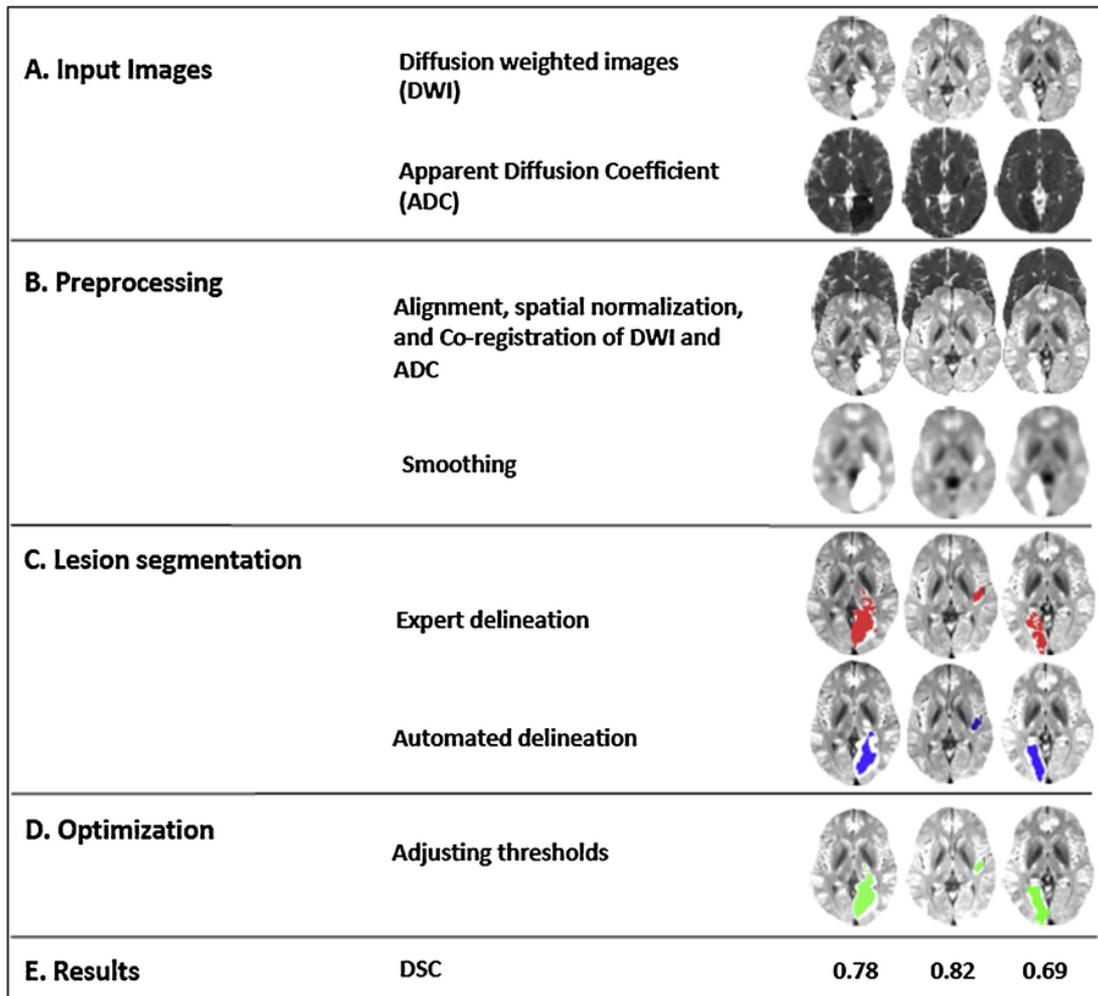


Fig. 2. The full segmentation algorithm with examples of DSC results on three AIS subjects. Panel (A) shows the input images (the DWI and the ADC images). Panel (B) shows the pre-processed variables, including alignment, spatial normalization, co-registration of ADC and DWI, and image smoothing. Panel (C) illustrates lesion segmentation delineated by experts (red regions) and statistical model (blue regions) overlaying the original DWI. Panel (D) shows the fine-tuned output by optimizing the thresholds. Finally panel (E) shows the segmentation performance by DSC.

Table 1

Overlap metrics for different configurations on train set with images smoothed by Gaussian filter with FWHM of 4 mm.

Configuration	Percentile of t-map on DWI (%)	Percentile of t-map on ADC (%)	Intensity of DWI (%)	DSC mean (SD)	Net Overlap mean (SD)
1	95	1	95	0.40 (0.19)	0.45 (0.18)
2	95	1	99	0.43 (0.18)	0.39 (0.19)
3	95	5	95	0.40 (0.22)	0.59 (0.20)
4	95	5	99	0.47 (0.21)	0.49 (0.20)
5	99	1	95	0.43 (0.17)	0.44 (0.19)
6	99	1	99	0.45 (0.18)	0.39 (0.19)
7	99	5	95	0.47 (0.21)	0.56 (0.19)
8	99	5	99	0.50 (0.19)	0.49 (0.20)

4. Discussion

We developed a fully automated lesion segmentation method, which relies on diffusion restriction characteristics of the acute stroke images. The methodology has proven to perform relatively well with the acute stroke cases with combinations of MRI modalities, namely DWI and ADC. The presented approach yielded good agreement with the expert delineated lesions, and high accuracy in classifying stroke images from non-stroke ones. One of the main advantages of our technique is that it mimics the current gold standard manual lesion delineation by considering both DWI and ADC images (rather than only using a fixed ADC threshold). Although our method does implement a thresholding step, it is important to note that the threshold is not fixed. Instead, it is population-derived approach which was fairly optimized throughout this study. Despite modest hardware requirements, the performance of our method is comparable with those obtained by GPU-dependent deep learning methods (Perez Malla et al., 2019; Subbanna et al., 2019, Clérigues et al., 2019, Boldsen et al., 2018), which may also be difficult to interpret.

The performance of our method is comparable with the previous studies, also with the state-of-the-art machine learning and/or deep learning techniques. Accordingly, (Perez Malla et al., 2019), reported a mean Dice similarity coefficient (DSC) of 0.34 by adapting a convolu-

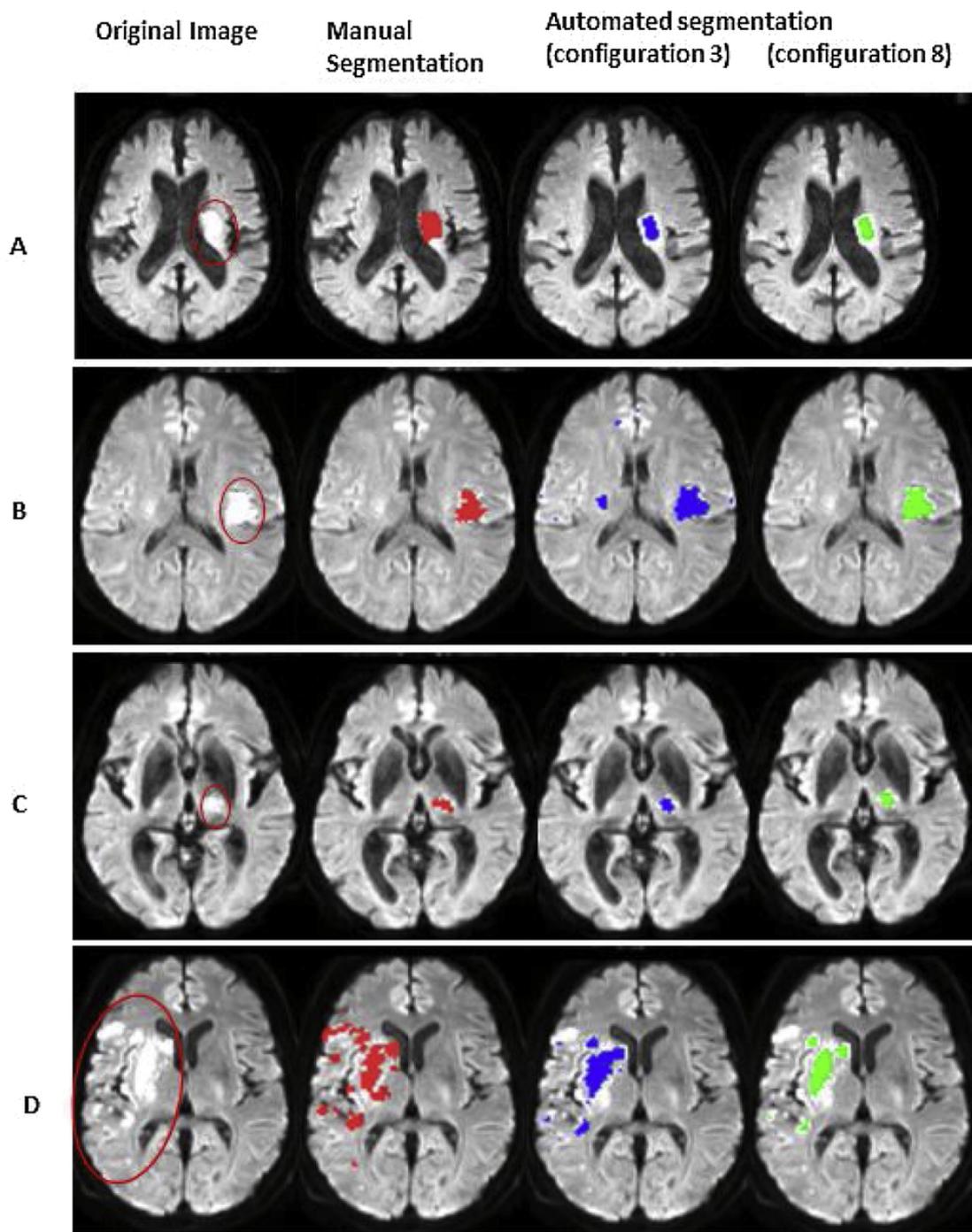


Fig. 3. Examples of the manual and automated lesion segmentations. The first column shows the original DWI and the second column shows the manual delineation of the acute lesions. The third and fourth columns demonstrate the results given by the proposed method with the third and eighth configurations, respectively. Panels A–C show cases where the algorithm performs accurately, whereas panel D shows a subject where the segmentation works suboptimally due to variable lesion size, and lesions are partially missed.

tional neural network (CNN) pipeline including transfer learning and data augmentation on 43 stroke patients, Subbanna et al. (2019) proposed a Markov random field model based approach which revealed an average DSC of 0.58 (tested on 151 multi-center datasets), and Boldsen et al. (2018) suggested an automatic tree learning segmentation and reported a median DSC of 0.61 on 108 patients. Recently, Cl'erigues et al. (2019) and Kamnitsas et al. (2017) reported an average DSC of 0.59 and 0.66 (tested only on 30 stroke cases from Ischemic Stroke Lesion Segmentation Challenge 2015) using a deep neural network frameworks including conditional random field post processing, respectively.

Overall, cascade of convolutional neural networks seem to yield better results for this purpose as, for example, (Chen et al., 2017) validated a lesion segmentation framework consists of two convolutional neural networks on DW images from 741 subjects which revealed a Dice coefficient of 0.67 in total. In a recent study, Winzeck et al. (2019) presented an ensemble of convolutional neural networks trained on multi-parametric DWI, ADC, and low b-value-weighted images from 116 acute stroke patients and achieved Dice scores up to 82.2 on a test set with 151 subject. Nonetheless, we believe direct and head-to-head comparison of these results would not be reasonable due to different datasets in terms of sample size and lesion characteristics.

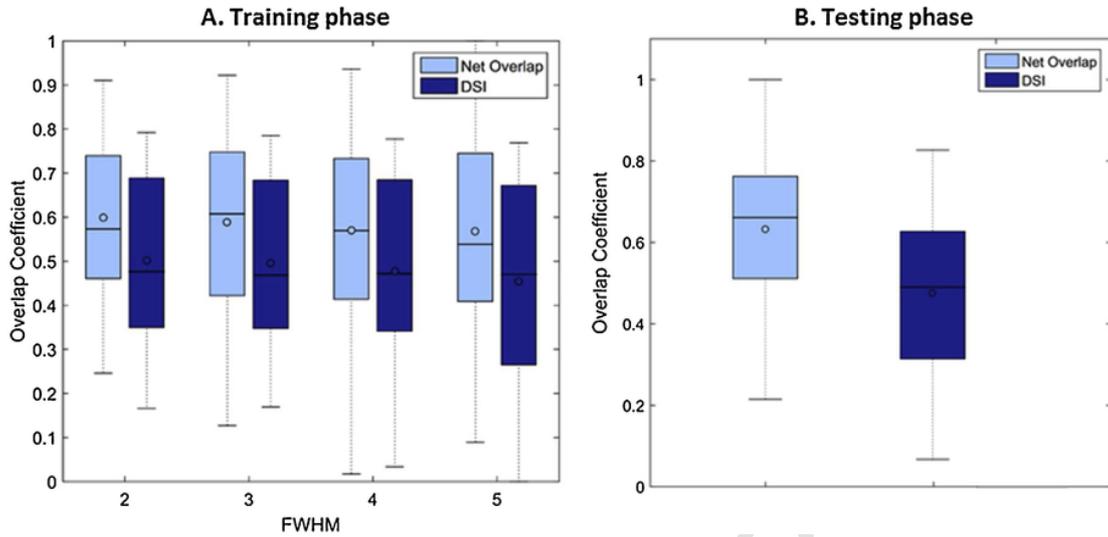


Fig. 4. Overlap metrics of the lesion segmentation method with the proposed configurations (eighth configuration for the overall overlap and third one for the net overlap) for Gaussian kernels with different FWHMs (from 2 to 5) on training set (A). Performance evaluation (net overlap and DSC) of the lesion segmentation method for the selected configuration and FWHM on test set (B). Error bars indicate standard deviations.

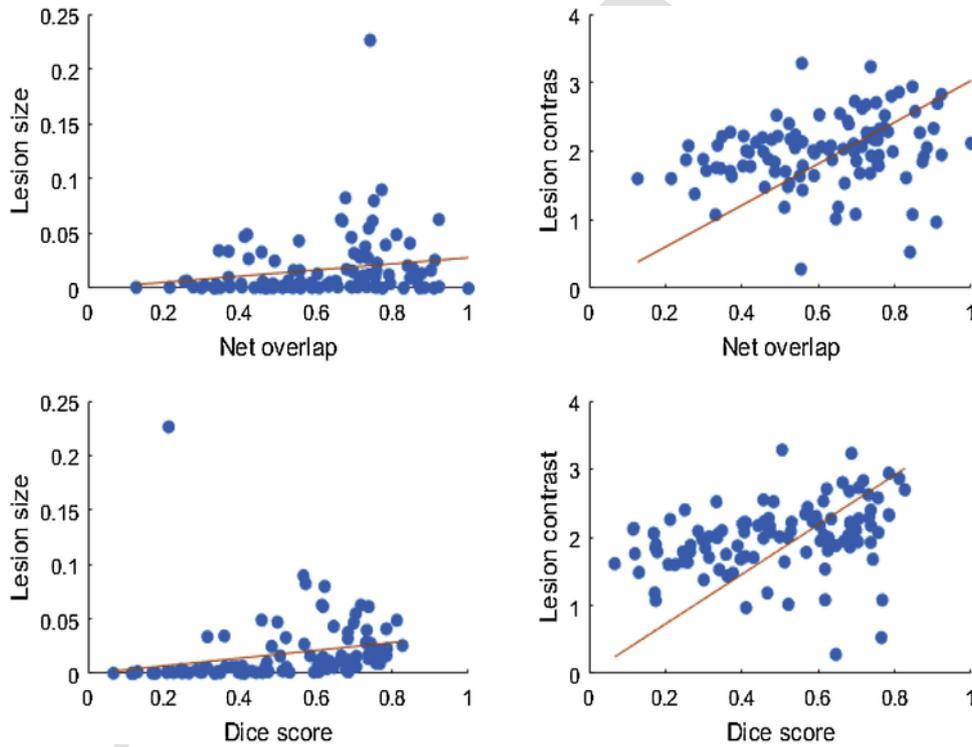


Fig. 5. Correlation test between spatial overlap metrics, size of the lesions, and intensity ratio. The experimental results for the classification of strokes from non-strokes showed mean accuracy, precision, sensitivity, and specificity of approximately 73 %, 77 %, 84 %, and 69 %.

We considered the impact of the image smoothing on the accuracy of lesion segmentation by applying Gaussian kernels with different FWHMs (from 2 to 5) to our images. This was done because smoothing the images with high FWHMs lowers the resolution and accordingly small lesions (only a few voxels large) might not be identified. On the other hand, the use of unsmoothed images or the ones smoothed with a very low FWHM is not recommended for the group analysis because of the variations in brain anatomy among individuals. Our results demonstrated that this method is robust to the FWHM used for smoothing. However, slight smoothing (FWHM of 3) resulted in best segmentation performance. This finding is close to what Gillebert and coworkers (Gillebert et al., 2014) found in 2014 (FWHM of 5), but contrary

to that of Stamatakis and Tyler in 2005 (Stamatakis and Tyler, 2005b) (FWHM of 8–12). However, as both studies commented, the size of the smoothing kernel should be contingent on the size of the lesions that need to be detected.

Our sample size was large and heterogeneous, with variable lesion location, size, shape, and contrast. Consequently, it is difficult to develop a method that works perfectly for all the cases. Accordingly, our approach yields slightly variable accuracy from case to case, as can be seen in Fig. 3(A–D). We found that lesion size and contrast affected the performance of the segmentation model. Our method performed better on larger lesions. This relationship may partly be explained by the smoothing of the images which decrease the resolution and makes

it more difficult to detect small lesions. This finding accords with previous observations (Gillebert et al., 2014; Perez Malla et al., 2019). Image contrast also showed a positive correlation with the performance of the method. Since this method works based on the detection of hyper and hypo intensities on the DWI and ADC images respectively, it was expected that sharper difference in intensities (contrast) between the lesion and healthy tissue lead to better performance of the method. Future work with significantly larger and high contrast datasets could test which lesion features are easiest to classify with the approach, allowing fine-tuning of the protocol to accommodate more lesion subtypes.

In this study, we measured the detection rate. This metric does not take the spatial overlap of the manual and automated masks into account. Instead, it examines if the method has been able to pinpoint the location of the lesion, though it could not delineate the boundaries perfectly. Due to the high detection rate (100 %) of our method, it can be used for automatic processing of DWIs in picture archiving and communication system (PACS) to alert the radiologist and clinicians if needed. This may yield significant savings in radiologists' workload, as potentially lesion-containing images can be automatically flagged for prioritized human validation. Furthermore, the method can be accommodated to the routine diagnosis pipeline of the hospitals for the initial screening followed by supervision of clinicians to refine the automatic segmentation. Finally, the method can be implemented in large-scale neuroinformatics studies where large number of images essentially precludes human delineation approach. From practical point of view, this is beneficial and could save time and energy for the clinicians and radiologist, as the algorithm can be used in conjunction with human operators for pinpointing the suspicious sites likely containing lesions.

4.1. Limitations and future directions

Certain limitations of the study should be acknowledged. First, the scope of this study was limited in terms of acute lesion segmentation. The reason is that in subacute lesions diffusion restrictions – seen as hyperintensity on DWI and hypointensity on ADC in the acute phase – are already resolved. Another limitation of this study is that our framework might be sensitive to DWI artifacts such as magnetic susceptibility artifacts at air-bone interfaces and/or from blood products resembling areas of restricted diffusion. In fact, these artifacts and infarcts pose similar imaging characteristics that can potentially lead to false detection of true lesions. However, this issue is not only occurring in intensity-based methods, such as current study, but also state-of-art neural network-based methods have reported similar limitation (Bhanu Prakash et al., 2006).

None of our optimization configurations were able to create lesion masks by which a generalized high DSC is achieved for all the study subjects. In fact, it was possible to individually optimize the configurations to obtain high DSC between the automated and manually delineated lesion masks, but having a fixed configuration that can lead to best DSC for all the subjects seemed impossible. This lack of generalizability might be due to the fact that MRI intensity values are usually considered unbounded.

Furthermore, we found a number of inaccurate segmentations in the border of the lesions, as shown in Fig. 3. This discrepancy, possibly due to the spatial normalization, smoothing of the images, and different contrast of the images, has been remarked in the earlier studies (Gillebert et al., 2014; Cl'rigues et al., 2019). Finally, manually segmented lesions by human experts, which served as the golden-standard in this study, may lead the model to include a systemic error in estimation of the lesion size (Ay et al., 2008; Campbell et al., 2010). Our automated segmentation method revealed small lesions in several cases in which manually drawn lesions failed to show (see Fig. 3, row D). However, as previously addressed by Boldsen et al. (2018), currently there is no substitute to circumvent this problem.

The method could be further improved. For example, providing more none-stroke images as a baseline may benefit the model decreasing the number of false positives due to the differences in individual's brain anatomy; therefore, resulting in a more stable model. Having more data for the model training might also reveal generalized and more optimal configuration. Second, to circumvent the difficulties associated with intensity-based methods, state-of-the-art methods (e.g. deep neural networks) could be applied to the data collected in this study. These methods mainly refer to the deep neural network (DNN) concept, which require a very large amount of data. In order to deal with the problem of data insufficiency for these methods, one solution would be data augmentation methods such as generative adversarial networks (GAN)(Yi et al., 2018). Another solution is applying transfer learning by reusing a model which is already trained by a large enough dataset (pre-trained models). Transfer learning empowers to train DNN models with relatively small datasets. An interesting future topic is to implement a multiclass classification set up to categorize stroke lesions into acute and sub-acute or even more precisely into four groups including focal hemorrhage, extended hemorrhage, focal ischemic and extended ischemic.

5. Conclusions

Triaging AIS patients for treatment requires efficient detection and segmentation of the AIS lesions. Manual lesion delineation on MRIs is expensive in terms of time and effort. Therefore, computer-assisted lesion segmentation methods that can accelerate or replace manual lesion delineation by experts are of interest. The current automated lesion segmentation method showed good agreement with the expert delineated lesions. The method is straightforward, fast, and does not require high computing power and memory. We believe this is advantageous and allows the technique to be implemented on an ordinary computer. However, further improvements and higher DSC are still desirable.

Authors' contributions

Authors' roles: Study design: Lauri Nummenmaa, Mikko Nyman, Sanaz Nazari-Farsani, and Tomi Karjalainen. Data collection: Mikko Nyman, Janne Isojärvi, and Marco Bucci. Data analysis: Sanaz Nazari-Farsani, Tomi Karjalainen, and Marco Bucci. Data interpretation: Sanaz Nazari-Farsani, Mikko Nyman, Tomi Karjalainen, Lauri Nummenmaa. Drafting manuscript: Sanaz Nazari-Farsani. Revising manuscript content: Sanaz Nazari-Farsani, Tomi Karjalainen, Marco Bucci, and Lauri Nummenmaa. Approving final version of manuscript: Sanaz Nazari-Farsani, Mikko Nyman, Tomi Karjalainen, Marco Bucci, Janne Isojärvi, and Lauri Nummenmaa. Sanaz Nazari-Farsani takes responsibility for the integrity of the data analysis.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We gratefully thank the Sigrid Juselius Foundation, Academy of Finland, State Funding for University-level Health Research, and TYKS Foundation for the financial support of this project. We would also like to thank Annastiina Mäkinen and Noora Nyssönen for the manual lesion segmentation on MR images of this study.

References

- Albers, G.W., Lansberg, M.G., Kemp, S., Tsai, J.P., Lavori, P., Christensen, S., Mlynash, M., Kim, S., Hamilton, S., Yeatts, S.D. 2017.

- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26, 839–851.
- Ay, H., Arsava, E.M., Vangel, M., Oner, B., Zhu, M., Wu, O., Singhal, A., Koroshetz, W.J., Sorensen, A.G., 2008. Interexaminer difference in infarct volume measurements on MRI: a source of variance in stroke research. *Stroke* 39, 1171–1176.
- Barber, P.A., Darby, D.G., Desmond, P.M., Yang, Q., Gerraty, R.P., Jolley, D., Donnan, G.A., Tress, B.M., Davis, S.M., 1998. Prediction of stroke outcome with echoplanar perfusion- and diffusion-weighted MRI. *Neurology* 51, 418–426.
- Bhanu Prakash, K., Gupta, V., Bilello, M., Beauchamp, N.J., Nowinski, W.L., 2006. Identification, Segmentation, and Image Property Study of Acute Infarcts in Diffusion-Weighted Images by Using a Probabilistic Neural Network and Adaptive Gaussian Mixture Model. *Acad. Radiol.* 13, 1474–1484.
- Boldsen, J.K., Engedal, T.S., Pedraza, S., Cho, T., Thomalla, G., Nighoghossian, N., Baron, J., Fiehler, J., Åstergaard, L., Mouridsen, K., 2018. Better diffusion segmentation in acute ischemic stroke through automatic tree learning anomaly segmentation. *Front. Neuroinform.* 12, 21.
- Campbell, B.C., Christensen, S., Foster, S.J., Desmond, P.M., Parsons, M.W., Butcher, K.S., Barber, P.A., Levi, C.R., Bladin, C.F., Donnan, G.A., Davis, S.M., 2010. EPITHET Investigators. Visual assessment of perfusion-diffusion mismatch is inadequate to select patients for thrombolysis. *Cerebrovasc. Dis.* 29, 592–596.
- Cawley, G.C., Talbot, N.L., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. 11:2079–107.
- Chen, L., Bentley, P., Rueckert, D., 2017. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *Neuroimage Clin.* 15, 633–643.
- Cl' erigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., Llad'o, X., 2019. Acute and Sub-Acute Stroke Lesion Segmentation From Multimodal MRI.
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory. 3:263–74.
- Crawford, J.R., Garthwaite, P.H., Howell, D.C., 2009. On comparing a single case with a control sample: an alternative perspective. *Neuropsychologia* 47, 2690–2695.
- Crawford, J.R., Garthwaite, P.H., Ryan, K., 2011. Comparing a single case to a control sample: Testing for neuropsychological deficits and dissociations in the presence of covariates. 47:1166–78.
- Dou, Q., Chen, H., Yu, L., et al., 2016. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans. Med. Imaging* 35, 1182–1195.
- Fiez, J.A., Damasio, H., Grabowski, T.J., 2000. Lesion segmentation and manual warping to a reference brain: intra- and interobserver reliability. *Hum. Brain Mapp.* 9, 192–211.
- Gillebert, C.R., Humphreys, G.W., Mantini, D., 2014. Automated delineation of stroke lesions using brain CT images. *Neuroimage Clin.* 4, 540–548.
- Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., Wolz, R., Valdés-Hernández, M.C., Dickie, D.A., Wardlaw, J., Rueckert, D., 2018. White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. 17:918–34.
- Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. 36:61–78.
- Legland, D., Kièu, K., Devaux, M., 2011. Computation of Minkowski measures on 2D and 3D binary images. 26:83–92.
- Mah, Y.H., Jager, R., Kennard, C., Husain, M., Nachev, P., 2014. A new method for automated high-dimensional lesion segmentation evaluated in vascular injury and applied to the human occipital lobe. *Cortex* 56, 51–63.
- Maier, O., Menze, B.H., von der Gablentz, J., Hani, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Gotz, M., Haeck, T., Halme, H.L., Havaei, M., Iftekaruddin, K.M., Jodoin, P.M., Kamnitsas, K., Kellner, E., Korvenoja, A., Larochele, H., Ledig, C., Lee, J.H., Maes, F., Mahmood, Q., Maier-Hein, K.H., McKinley, R., Muschelli, J., Pal, C., Pei, L., Rangarajan, J.R., Reza, S.M.S., Robben, D., Rueckert, D., Salli, E., Suetens, P., Wang, C.W., Wilms, M., Kirschke, J.S., Kr Amer, U.M., Munte, T.F., Schramm, P., Wiest, R., Handels, H., Reyes, M., 2015. ISLES 2015 - a Public Evaluation Benchmark for Ischemic Stroke Lesion Segmentation From Multispectral MRI.
- Maier, O., Schroder, C., Forkert, N.D., Martinetz, T., Handels, H., 2015. Classifiers for ischemic stroke lesion segmentation: a comparison study. *PLoS One* 10, 0145118.
- Martel, A.L., Allder, S.J., Delay, G.S., Morgan, P.S., Moody, A.R., 1999. Measurement of Infarct Volume in Stroke Patients Using Adaptive Segmentation of Diffusion Weighted MR Images.: 22–31.
- Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., 2014. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, 1993–2024.
- Mohd Saad, N., Abu-Bakar, S.A.R., Muda, S., Mokji, M.M., Salahuddin, L., 2011. Brain lesion segmentation of diffusion-weighted MRI using gray level co-occurrence matrix. In: 2011 IEEE International Conference on Imaging Systems and Techniques. pp. 284–289.
- Neumann, A.B., Jonsdottir, K.Y., Mouridsen, K., Hjort, N., Gyldensted, C., Bizzi, A., Fiehler, J., Gasparotti, R., Gillard, J.H., Hermier, M., Kucinski, T., Larsson, E.M., Sorensen, L., Ostergaard, L., 2009. Interrater agreement for final infarct MRI lesion delineation. *Stroke* 40, 3768–3771.
- Nogueira, R.G., Jadhav, A.P., Haussen, D.C., Bonafe, A., Budzik, R.F., Bhuva, P., Yavagal, D.R., Ribo, M., Cognard, C., Hanel, R.A., 2018. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N. Engl. J. Med.* 378, 11–21.
- Perez Malla, C.U., Valdes Hernandez, M.D.C., Rachmadi, M.F., Komura, T., 2019. Evaluation of enhanced learning techniques for segmenting ischaemic stroke lesions in brain magnetic resonance perfusion images using a convolutional neural network scheme. *Front. Neuroinform.* 101477957 PMC.
- Powers, W.J., Rabinstein, A.A., Ackerson, T., Adeoye, O.M., Bambakidis, N.C., Becker, K., Biller, J., Brown, M., Demaerschalk, B.M., Hoh, B., 2018. 2018 guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 49, 46–99.
- Stamatakis, E.A., Tyler, L.K., 2005. Identifying lesions on structural brain images—validation of the method and application to neuropsychological patients. *Brain Lang.* 94, 167–177.
- Stamatakis, E.A., Tyler, L.K., 2005. Identifying lesions on structural brain images—validation of the method and application to neuropsychological patients. *Brain Lang.* 94, 167–177.
- Straka, M., Albers, G.W., Bammer, R., 2010. Real-time diffusion-perfusion mismatch analysis in acute stroke. 32:1024–37.
- Subbanna, N.K., Rajashekar, D., Cheng, B., Thomalla, G., Fiehler, J., Arbel, T., Forkert, N.D., 2019. Stroke Lesion Segmentation in FLAIR MRI Datasets Using Customized Markov Random Fields. 10:541.
- Tsai, J.Z., Peng, S.J., Chen, Y.W., Wang, K.W., Wu, H.K., Lin, Y.Y., Lee, Y.Y., Chen, C.J., Lin, H.J., Smith, E.E., Yeh, P.S., Hsin, Y.L., 2014. Automatic detection and quantification of acute cerebral infarct by fuzzy clustering and histogram characterization on diffusion weighted MR imaging and apparent diffusion coefficient map. *Biomed Res. Int.* 2014, 963032.
- Vert, C., Parra-Fariñas, C., Rovira, , 2017. MR imaging in hyperacute ischemic stroke. *Eur. J. Radiol.* 96, 125–132.
- Winzeck, S., Mocking, S.J.T., Bezerra, R., MJRJ, Bouts, McIntosh, E.C., Diwan, I., Garg, P., Chutinet, A., Kimberly, W.T., Copen, W.A., Schaefer, P.W., Ay, H., Singhal, A.B., Kamnitsas, K., Glocker, B., Sorensen, A.G., Wu, O., 2019. Ensemble of convolutional neural networks improves automated segmentation of acute ischemic lesions using multiparametric diffusion-weighted MRI. *Am. J. Neuroradiol.*
- Yi, X., Wallia, E., Babyn, P., 2018. Generative Adversarial Network in Medical Imaging: A Review, abs/1809.07294.
- Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells 3rd, W.M., Jolesz, F.A., Kikinis, R., 2004. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* 11, 178–189.