



OPEN

DATA DESCRIPTOR

ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset

Moritz R. Hernandez Petzsche¹✉, Ezequiel de la Rosa^{2,3}, Uta Hanning⁴, Roland Wiest⁵, Waldo Valenzuela⁵, Mauricio Reyes⁶, Maria Meyer², Sook-Lei Liew⁷, Florian Kofler^{1,3,8,9}, Ivan Ezhov^{3,8}, David Robben², Alexandre Hutton⁷, Tassilo Friedrich¹, Teresa Zarth¹, Johannes Bürkle¹, The Anh Baran¹, Björn Menze^{3,10}, Gabriel Broocks⁴, Lukas Meyer⁴, Claus Zimmer¹, Tobias Boeckh-Behrens¹, Maria Berndt¹, Benno Ikenberg¹¹, Benedikt Wiestler^{1,8} & Jan S. Kirschke^{1,8}

Magnetic resonance imaging (MRI) is an important imaging modality in stroke. Computer based automated medical image processing is increasingly finding its way into clinical routine. The Ischemic Stroke Lesion Segmentation (ISLES) challenge is a continuous effort to develop and identify benchmark methods for acute and sub-acute ischemic stroke lesion segmentation. Here we introduce an expert-annotated, multicenter MRI dataset for segmentation of acute to subacute stroke lesions (<https://doi.org/10.5281/zenodo.7153326>). This dataset comprises 400 multi-vendor MRI cases with high variability in stroke lesion size, quantity and location. It is split into a training dataset of $n = 250$ and a test dataset of $n = 150$. All training data is publicly available. The test dataset will be used for model validation only and will not be released to the public. This dataset serves as the foundation of the ISLES 2022 challenge (<https://www.isles-challenge.org/>) with the goal of finding algorithmic methods to enable the development and benchmarking of automatic, robust and accurate segmentation methods for ischemic stroke.

Background & Summary

Stroke is a leading cause of morbidity and mortality worldwide¹. Up to two thirds of stroke survivors suffer permanent disability². In the last decade, the advent of endovascular reperfusion therapy has significantly improved stroke outcome in patients with large vessel occlusions^{3–6}. Image-based guidance of revascularization treatment decisions has further improved patient outcome for computer tomography (CT)^{7–9} and magnetic resonance imaging (MRI)^{10,11}. Computer aided image analysis, especially for CT perfusion data has already found entry into clinical routine in many centers and is recommended by national guidelines¹² to aid decision making regarding reperfusion therapy^{13–16}. Machine learning and deep learning approaches have been shown to facilitate clinical interpretation of CT perfusion data and have been widely adopted in clinical routine^{17–22}. Segmentation based volumetric analyses of stroke lesions in magnetic resonance imaging (MRI) are often performed for research purposes and have been shown to predict clinical outcome^{23–26}. However, stroke lesion segmentations are usually painstakingly performed by hand and the quality of annotations are heavily dependent

¹Department of Diagnostic and Interventional Neuroradiology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. ²icomatrix, Leuven, Belgium. ³Department of Informatics, Technical University of Munich, Munich, Germany. ⁴Department of Diagnostic and Interventional Neuroradiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ⁵Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Bern, Switzerland. ⁶ARTORG Center for Biomedical Engineering Research, Univ. of Bern, Bern, Switzerland. ⁷Chan Division of Occupational Science and Occupational Therapy, University of Southern California, Los Angeles, CA, USA. ⁸TranslaTUM – Central Institute for Translational Cancer Research, Technical University of Munich, Munich, Germany. ⁹Helmholtz AI, Helmholtz Zentrum Munich, Munich, Germany. ¹⁰Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland. ¹¹Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany. ✉e-mail: moritz.hernandez@tum.de

on preexisting neuroimaging experience of the rater and the total time and effort invested. The time-consuming nature of this task prevents the regular use of segmentations during clinical routine, which is further impeded by a high inter-observer variability. Automated annotation of stroke lesions could be used in clinical routine to guide therapeutic decisions in an acute setting and to predict outcome at the subacute to chronic stage. Stroke lesion segmentation could also be used to automatically classify stroke etiology in post-stroke MRI.

The first Ischemic Stroke Lesion Segmentation (ISLES) challenge, which took place in 2015, was split into two sub-challenges: Sub-acute Stroke Lesion Segmentation (SISS) and Stroke Perfusion Estimation (SPES). The goal of SISS (with a total of 64 cases for training and testing) was to segment subacute stroke lesions using conventional post-stroke MRI sequences, including T2 and T1 weighted imaging, fluid attenuated inversion recover (FLAIR), and DWI²⁷. The ISLES challenge 2018, which was the previous challenge edition, was set up to predict infarct core delineated in diffusion weighted imaging (DWI) using CT perfusion data²¹. Both ISLES events received major attention from the research community: there were 120 database downloads until the ISLES15 challenge day with 14 participating teams, and the number of participating teams was roughly duplicated in the latest ISLES'18 edition. The ISLES'15 and ISLES'18 challenges played a crucial role in identifying prominent methods for acute and sub-acute ischemic stroke lesion segmentation. These datasets have since served as important benchmarks for the scientific community.

Based on the experience gained from these previous editions, ISLES'22 aims to benchmark acute and sub-acute ischemic stroke MRI segmentation using 400 cases, a more than 6-fold increase in numbers compared to ISLES15. This dataset is provided to train and benchmark DWI infarct segmentation in acute and sub-acute stroke. ISLES'22 differs in several ways from the previous challenge editions in ischemic stroke by: 1) targeting the delineation of not only large infarct lesions, but also of multiple embolic and/or cortical infarcts (typically seen after mechanical recanalization), 2) by evaluating both pre- and post- interventional MRI images in a multicenter and multi-scanner dataset, 3) generalizability of the models will be tested in a hidden dataset which also includes cases from a center that the algorithms were not exposed to in the training stage. For detailed information about the ISLES'22 challenge event, readers are referred to²⁸. This challenge, together with the ATLAS challenge (<https://atlas.grand-challenge.org/>)²⁹ are using the web-based platforms <http://www.isles-challenge.org/> and <https://isles22.grand-challenge.org/>.

We chose to include all imaging sequences relevant for the radiological diagnosis of acute to subacute stroke lesions in MRI (FLAIR, DWI and ADC). The information contained in these imaging sequences allow a secure diagnosis of infarct and differentiation of infarcts from other cerebral lesions like older glial scars (gliosis, this may be due to a range of differing pathologies) or artefacts in DWI (e.g. skull base artefacts). The inclusion of all three imaging parameters also allow a precise spatial differentiation of infarct vs. healthy brain tissue. In the parallel held ATLAS challenge²⁹, segmentation is performed solely on T1 in the chronic stage of stroke, where infarcts are reduced to a glial scar. In this setting, it is much more difficult to accurately segment the infarct borders and to visualize smaller infarcts. Additionally, it is difficult to accurately determine the etiology of the gliosis (e.g. infarct vs. trauma or bleeding).

In the challenge for the here described dataset, teams will deal with a wider ischemic stroke disease spectrum, involving variable lesion size and burden, complex infarct patterns and variable anatomical lesion location in data from multiple centers. The diversity of the ISLES'22 dataset will provide a unique challenge for participants.

Methods

Ethical statement. This retrospective evaluation of imaging data was approved by the local ethics boards of all participating centers. Requirement of written informed consent was waived by all ethics boards due to the retrospective nature of the study and the rigorous patient de-identification of the data.

Subject selection. Inclusion criteria for the dataset: Subjects 18 years or older who had received MR imaging of the brain for previously diagnosed or suspected stroke were included in this study. The imaging protocol required at least a FLAIR and DWI sequence. DWI consists of a trace image at a b-value up to 1000 s/mm² as well as its corresponding apparent diffusion coefficient (ADC) map. Image acquisition was performed on one of the following devices: 3 T Philips MRI scanners (Achieva, Ingenia), 3 T Siemens MRI scanner (Verio) or 1.5 T Siemens MAGNETOM MRI scanners (Avanto, Aera). MRIs included were intentionally chosen to be heterogeneous to ensure the best possible training and generalization of the algorithms. See Table 1 for an overview of MRI acquisition parameters by center. Images were obtained by healthcare professionals as part of the clinical imaging routine for stroke patients at three different stroke centers.

Care was taken to select a broad spectrum of infarct patterns. All vascular territories were included at a similar rate. A large subset of patients with posterior circulation infarct (e.g. due to basilar artery occlusion) was included in this study. Due to a high degree of closely situated skull-base artefacts in DWI, infratentorial infarcts are difficult to segment for unexperienced raters and are frequently overlooked by segmentation algorithms with little training exposure to posterior circulation infarct. These additional labeling challenges lead us to include more cases of posterior circulation ischemia than would be expected if case selection were truly random. Figure 1 shows sample cases portraying the infarct spectrum included in this dataset.

Segmentation difficulties also frequently arise in cases with large amounts of punctiform infarcts. For human raters, to adequately capture the entire infarct territory, a time-consuming segmentation effort is required. Similarly, machine trained algorithms frequently fail to capture all the small affected regions. In $n = 3$ patients in the training dataset, where MR imaging was acquired for suspected stroke, no infarct was found. We chose to include this small subset to diversify the dataset. Table 2 gives an overview of infarct volumes per scan (scan infarct volume) and per unconnected lesion (lesion-wise infarct volume) for each center, as well as the number of unconnected infarcts per scan in the entire dataset. The number of unconnected infarcts was calculated using the python library cc3d³⁰.

Center	Modality	TR (ms)	TE (ms)	TI (ms)	Flip Angle (°)	Voxel (mm ²)	Slice thickness (mm)	Gap (mm)
#1	FLAIR	[4800–12000]	[120–395]	[1650–2850]	[90–180]	[0.23 × 0.23–1.00 × 1.00]	[0.68–5.00]	[0.71–5.00]
	DWI	[3175–16439]	[55–91]	—	[90–90]	[0.87 × 0.87–2.00 × 2.00]	[2.00–5.00]	[2.00–5.00]
#2	FLAIR	[8500–8500]	[82–111]	[2440–2440]	[150–150]	[0.43 × 0.43–0.86 × 0.86]	[4.80–9.60]	[4.80–6.00]
	DWI	[3500–4780]	[67–89]	—	[90–180]	[0.57 × 0.57–2.00 × 2.00]	[3.00–6.50]	[4.80–4.80]
#3	FLAIR	[6500–8000]	[103–104]	[2200–2370]	[150–150]	[0.45 × 0.45–0.45 × 0.45]	[6.50–6.50]	[6.5–6.5]
	DWI	[13000–13000]	[90–90]	—	[90–90]	[0.89 × 0.89–1.87 × 1.87]	[2.00–6.50]	[0.00–0.00]
All	FLAIR	[4800–12000]	[82–395]	[1650–2850]	[90–180]	[0.23 × 0.23–1.00 × 1.00]	[0.68–9.60]	[0.71–6.50]
	DWI	[3175–16439]	[55–91]	—	[90–180]	[0.57 × 0.57–2.00 × 2.00]	[2.00–6.50]	[2.00–5.00]

Table 1. Overview of MR imaging parameters.

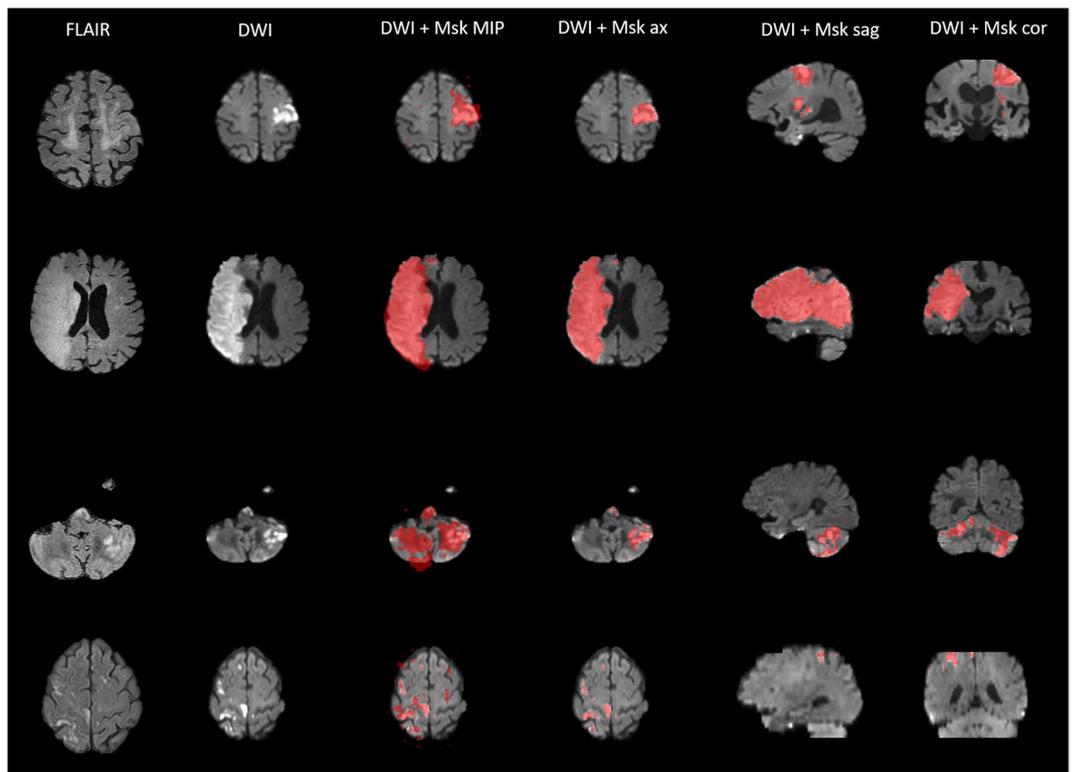


Fig. 1 Exemplary 3D Snapshots through the infarct center of mass. Axial FLAIR and DWI images are displayed in the two leftmost columns. The 3rd column from the left shows the maximum intensity projection (MIP) of the mask (Msk). The three rightmost columns show DWI with mask overlay (without MIP) in the three anatomical planes. Top row: Larger left-sided infarct with bilateral punctiform, likely embolic satellite ischemias. 2nd row: Infarct of the entire right sided middle cerebral artery territory. 3rd row: Bilateral cerebellar and occipital infarcts after posterior circulation ischemia. Bottom row: Bilateral punctiform infarcts, likely resulting from multiple micro-embolic occlusions.

In the hyper-acute phase of ischemic stroke, up to 4.5 hours post onset, restricted diffusion is present (high signal on DWI and low signal on ADC) while the FLAIR in the affected parenchyma may remain without changes. This imaging phenomenon is called a FLAIR-DWI mismatch and is used in clinical practice to estimate the time window in patients where the time of onset is unknown. An accurate estimate of the time of onset is crucial to make decisions regarding revascularization treatment¹¹. In acute ischemic stroke, following this hyper-acute phase and usually defined in literature as 0 to 7 days from onset, DWI and FLAIR show a high signal with reduced ADC values in the affected brain parenchyma. In the subacute stage, between 1 to 3 weeks post onset, high DWI signal begins to diminish while ADC first normalizes to values of healthy brain tissue, a phenomenon frequently referred to as pseudonormalization. FLAIR signal remains high. In the chronic stage, beginning 3 weeks after onset, DWI signal is variable but usually iso- to hypointense depending on underlying T2 signal, while ADC values are high^{31–40}. MRIs with late acquisition post stroke (>1 week) often lead to a decreased DWI signal intensity for ischemic brain parenchyma. A lower signal intensity in DWI leads to lower MRI sensitivity for stroke and segmentation difficulty for both human and machine raters. In these cases, it

Center	Statistic	Scan infarct volume (ml)	Number of unconnected infarcts	Lesion-wise infarct volume (ml)
#1	Mean (std)	27.15 (51.79)	8.21 (8.33)	3.31 (19.39)
	[Min, Max]	[0.00, 482.15]	[0, 46]	[0.01, 477.26]
#2	Mean (std)	17.45 (24.55)	11.57 (19.10)	1.51 (8.30)
	[Min, Max]	[0.00, 163.97]	[0, 126]	[0.01, 163.90]
#3	Mean (std)	40.77 (46.80)	8.58 (7.62)	4.75 (20.12)
	[Min, Max]	[0.23, 233.85]	[1, 37]	[0.01, 233.62]
All	Mean (std)	26.38 (46.23)	9.11 (12.06)	2.90 (16.77)
	[Min, Max]	[0.00, 482.15]	[0, 126]	[0.01, 477.26]

Table 2. Summary infarct lesion statistics for the ISLES 2022 dataset. Min: minimum; Max: maximum; std: standard deviation.

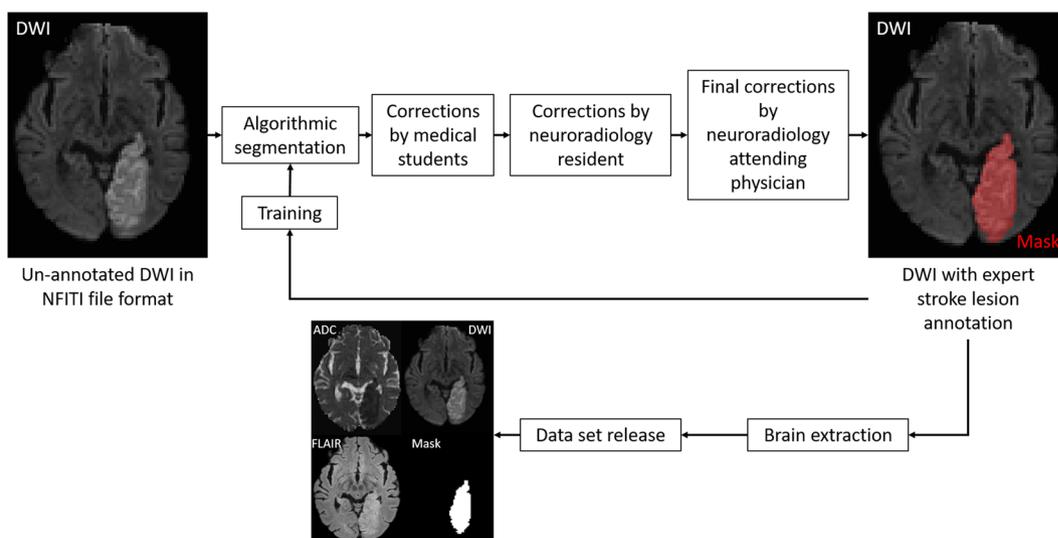


Fig. 2 Workflow graphic of the hybrid human-algorithm stroke lesion segmentation applied in this dataset.

is especially difficult to adequately annotate the border between infarct and healthy brain tissue. This dataset includes cases of MRIs in various stages of sub-acute stroke from multiple previous studies^{41–44} to find machine learning solutions to this frequent issue in stroke lesion segmentation.

Ground truth stroke lesion segmentation. A hybrid human-algorithm annotation scheme was applied to segment all cases from Center #1 and #3 (data from the Technical University Munich and University Medical Center Hamburg-Eppendorf, see below). First, the MR input data was anonymized by conversion to Neuroimaging Informatics Technology Initiative (NIFTI) format (<https://nifti.nimh.nih.gov/nifti-1/>), according to the Brain Imaging Data Structure (BIDS) convention (<https://bids.neuroimaging.io/>). As part of pre-processing, DWI and its corresponding ADC map were resliced using ANTs⁴⁵ to an axial isotropic voxel size of $2 \times 2 \text{ mm}^2$. This was performed only for the cases of Center #1 (see below), as this represents the original acquisition voxel size. The slice thickness remained as acquired. FLAIR data remained as exported.

In most cases for stroke lesion segmentation, it is easier to edit and revise an existing annotation than to create an annotation ‘from scratch’. Therefore, under consideration of the caseload in this challenge and in contrast to previous iterations of the ISLES challenge, a pre-segmentation algorithm (3D UNet⁴⁶) was trained on DWI data stemming from the Technical University Munich (Center #1, see below) that was previously annotated for other research projects. This algorithm was trained solely using a single MRI modality ($B = 1000 \text{ DWI}$). Manually pre-segmented data intended for the training of this algorithm underwent rigorous quality control. Annotations with suboptimal quality were frequent; these cases were edited by a neuroradiology resident and checked by a senior neuroradiologists before being admitted to the training of our pre-segmentation algorithm. These cases were also included in the ISLES challenge.

This house-trained algorithm later pre-segmented all thus far un-annotated scans intended for later release. These algorithmic segmentations were then checked and edited by medical students with special stroke lesion segmentation training. The pre-segmentation algorithm was updated once as more high-quality segmentations became available for training. This resulted in more accurate predictions and a lesser effort of correcting annotations by medical students. All medical-student edited annotations were critically revised and further edited by the neuroradiologist in training and the final data sets were reviewed and approved by one out of three

	ISLES vs Rater II		ISLES vs Rater III		Rater II vs Rater III	
	Dice	VD (ml)	Dice	VD (ml)	Dice	VD (ml)
Mean (std)	0.90 (0.09)	2.37 (2.59)	0.86 (0.13)	6.56 (13.37)	0.83 (0.12)	7.22 (11.32)
[Min, Max]	[0.69, 0.98]	[0.11, 7.34]	[0.61, 0.98]	[0.06, 45.66]	[0.57, 0.98]	[0.06, 38.32]

Table 3. Inter-rater analysis. VD: Volume difference; Dice: Dice coefficient; Min: minimum; Max: maximum; std: standard deviation.

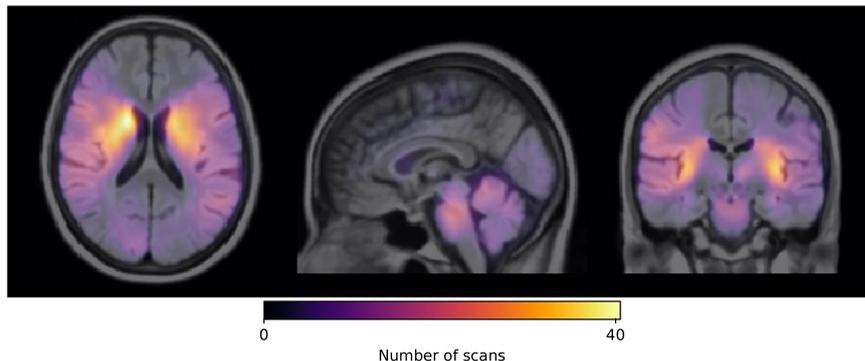


Fig. 3 ISLES'22 probabilistic lesion map overlapped to an MNI FLAIR template.

Dataset	Vascular territory			
	ACA	MCA	PCA	Infratentorial vasculature Pons / Medulla / Cerebellum
Train	15 (6.1%)	129 (52.2%)	26 (10.5%)	77 (31.2%)
Test	6 (4.1%)	93 (62.8%)	18 (12.2%)	31 (20.9%)
All	21 (5.3%)	222 (56.2%)	44 (11.1%)	107 (27.1%)

Table 4. ISLES'22 spatial lesion distribution across vascular territories. The number of scans (percentage) is shown. ACA: Anterior cerebral arteries; MCA: middle cerebral arteries; PCA: posterior cerebral arteries.

attending neuroradiologists, all of them with more than 10 years of experience in stroke imaging. If algorithmic segmentation was deemed insufficient, which sometimes occurred in the first version of the pre-segmentation algorithm, the medical students could discard the algorithmically generated mask and manually annotate 'from scratch'. The later instances of quality control (resident and senior physicians) were blinded to the method of pre-segmentation (algorithmic followed by student changes vs. 'from scratch' student segmentation), lowering the possibility of a resulting annotation bias. Manual stroke lesion segmentations were performed using ITK-Snap⁴⁷ or 3D Slicer⁴⁸ both open-source tools for brain imaging visualization and segmentation. Figure 2 shows an overview of annotation work-flow in preparation of the release of the dataset.

Although the pre-segmentation algorithm was trained only using a single MRI modality ($B = 1000$ DWI) for reasons of convenience, the final expert annotation includes in the information of all three later released modalities (DWI, ADC, FLAIR). Stroke lesion identification in clinical practice is always performed under consideration of all three released modalities. DWI is reviewed as a primary stroke imaging sequence. If the reader is unsure of the authenticity of a stroke lesion (e.g. due to suspected DWI artefact at the skull base or to differentiate T2 'shine-through' from true diffusion restriction), ADC and FLAIR are reviewed. All three sequences are therefore essential for accurate classification of stroke lesions. As in clinical practice, the expert raters annotating this dataset reviewed ADC and FLAIR in addition to DWI in order to achieve the best possible annotation. For this reason, readers aiming to use this dataset for the generation of segmentation algorithms are encouraged to integrate all three sequences into their approach in order to achieve the best possible result.

Post-annotation data pre-processing. In preparation for data-set release and in accordance to the ethical approval obtained for this challenge, the imaging data was irreversibly anonymized. For this de-identification of patients, brain-extraction was performed mask-based using the HD-BET algorithm⁴⁹ completion of the annotation process. FLAIR to DWI rigid registration was performed using Elastix^{50,51} and subsequent skull-stripping of DWI and ADC using the registered brain mask was performed. After skull stripping for de-identification, all imaging sequences were returned to their native space before data release.

Inter-rater analysis. In order to understand the impact of different expert annotations over the segmented stroke lesions, an inter-rater delineation experiment was conducted. In this experiment, 10 cases from ISLES'22

were selected and re-delineated by two expert neuroradiologists with more than 16 (rater II, JSK) and 10 (rater III, BW) years of experience in the field. The cases were chosen as variable as possible, including large and small infarcts, large vessel occlusion strokes and embolic pattern lesions, located in different anatomical brain areas. Dice coefficient was computed as a metric of delineation overlaps across raters. The lesion volume differences across raters were also included. The inter-rater results are summarized in Table 3.

In reviewing annotations that were previously performed for other research projects (these were used for primary training of our pre-segmentation algorithm after expert edits as described above), generally inadequate quality of the segmentations was a frequent finding. Most frequently, the expert raters of this dataset found that the infarct border was frequently traced too conservatively, leading to underestimation of the infarct volume. DWI decreases gradually in signal along the border of the infarct due to partial volume effects and thus, in order to accurately segment the edges, it is often helpful to review the high resolution FLAIR image, where infarct edema can be visualized and healthy tissue can be sharply differentiated from the infarct. Another frequent issue with segmentation, as discussed above, is the massively time-consuming effort required to accurately annotate a large number of punctiform infarcts resulting from an embolic shower. As this infarct pattern results frequently in patients who underwent successful mechanical thrombectomy, it frequently occurred within this dataset. Among the preexisting segmentations, we found that large quantities of small embolic infarcts were almost always inadequately annotated, further highlighting the need for an accurate algorithmic solution to stroke lesion segmentation.

The ground truth in this dataset, just like in any human-annotated dataset, is only as good as the performance of the human raters responsible for the annotation. With a special focus on the difficulties mentioned above, annotations were performed using a multi-layered human-algorithm hybrid workflow to distribute the work-load amongst the raters and to maximize overall segmentation accuracy. As is evident in Table 3, the annotations within the released dataset had higher Dice score and a lower volume difference with the individual expert raters than the expert raters amongst themselves.

Probabilistic lesion map of ISLES'22. A 3D, probabilistic lesion map was generated in order to understand the voxel-wise lesion distribution of the entire dataset (see Fig. 3). With this aim, all DWI scans from the ISLES'22 train and test datasets were linearly registered to a FLAIR MNI template (Schirmer *et al.*)⁵² using NiftyReg (Ourselin *et al.*)⁵³. Suboptimally registered images were identified through visual quality control and were later re-registered until achieving an appropriate alignment with the atlas. Later on, the lesions masks were projected to the MNI space using the DWI to FLAIR found registration matrices. The probabilistic atlas was finally obtained by counting, voxelwise, the number of scans that showed overlapped infarcted tissue. The 3D probabilistic lesion map is available in Nifty format through Zenodo (<https://zenodo.org/record/7335305>)⁵⁴. Sample snapshots of the probabilistic lesion map registered to an MNI FLAIR template are shown in Fig. 3.

Furthermore, we have quantified the spatial distribution of the stroke lesions according to the vascular territories irrigated by the middle, anterior, posterior cerebral arteries as well as by the infratentorial vasculature (including the pons, medulla and cerebellum areas). The spatial distribution of the lesions was computed by calculating the center of mass of the MNI-registered stroke lesions and by overlapping them to a vascular territory atlas (Schirmer *et al.*⁵², <https://doi.org/10.5281/zenodo.7335305>)⁵⁴, see Table 4.

Challenge metrics. The evaluation metrics for results submitted by competing teams in the ISLES challenge were chosen to evaluate performance of 1) correct segmentation of large lesions, especially their borders and 2) detection of most, if not all small punctiform infarcts. Therefore, individual segmentation metrics (as Dice coefficient only) are unlikely to be sufficient, since small lesions may not consistently drive changes in some overlap measures (for instance, in the presence of a large stroke lesion and a very small separated embolic infarct, a large Dice increase will come by only detecting the large lesion, even if the small lesion is missed). Thus, with a focus on clinical translation, we consider metrics that are often of main interest for neuroradiologists, such as the lesion volume, the presence/absence of a lesion (i.e. detection) and the accurate count of the lesion burden. However, we also included classical segmentation metric as Dice similarity coefficient to gain an overall impression of the performance overlap between ground truth and predictions. The following error metrics will be used for scoring:

Dice similarity coefficients for segmentation masks and the absolute difference for infarct volume will be computed as voxel-wise surrogates for model performance. As a lesion wise metric the lesion count absolute difference in the predicted mask will be calculated obtained by computing the amount of connected components per case. As an additional lesion-wise metric, the lesion detection f1 score will be calculated.

As in previous iterations of the ISLES challenge, ranking is produced by comparing each metric at the case level. In short, metrics are calculated for each case, followed by establishing metric-specific rank separately for each dataset. A mean rank over all metrics is obtained to obtain the team's rank for each case. The final rank is the mean of all case-specific ranks. Scoring will be performed on a script-automated basis.

Data Records

Data repository and storage. All training data (n = 250) has been made publicly available under the creative commons license CC-BY-4.0 in the preprocessed format on Zenodo (<https://doi.org/10.5281/zenodo.7153326>)⁵⁵. Further information about the ISLES challenges can be found under <http://www.isles-challenge.org/> and under <https://isles22.grand-challenge.org/>.

Data structure and file formats. All medical imaging files were exported from the Picture Archiving and Communication System in the NIfTI format. Segmentation masks are created and saved in NIfTI format. All data in the ISLES'22 dataset was separated into a training dataset (250 subjects) and a test dataset (150 subjects).

Corresponding scanner metadata from the Digital Imaging and Communications in Medicine (DICOM) header in the JSON file format is provided with the datasets, if available. MRIs from the following centers were included:

Center #1: University Hospital of the Technical University Munich, Munich, Germany.

Center #2: University Hospital of Bern, Bern, Switzerland.

Center #3: University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

The publicly available train set comprises data from centers #1 and #2. The test set comprises data from all the three centers in equal parts as follows:

Acute to early sub-acute stroke data from centers #1 and #3 (MRIs acquired after revascularization therapy).

Hyper-acute to acute stroke data from center #2 (MRIs acquired before revascularization therapy).

Thus, in this ISLES'22 task we will evaluate the robustness and generalization capability of the proposed models over 1) new scans coming from two centers already used at the training stage, 2) new scans coming from a new (unseen at training stage) center, and 3) new scans acquired before revascularization therapy, coming from a center already seen at training stage. Introduction of a previously unseen center at the test stage likely results in an overall deterioration of algorithmic performance but in turn allowing a better differentiation of the generalizability of the solutions.

The split between training and test data set has been performed so that both sets include a similar variance of stroke lesion patterns ranging from large territorial infarcts to small punctiform ischemia. The spatial distribution of lesions according to the different vascular territories is summarized in Table 4. The majority of the scans have lesions within the territories irrigated by the middle cerebral arteries and the infratentorial vasculature. Similar spatial distributions between training and test datasets have been considered. A slight difference in the distribution of infratentorial scans (~10% more cases in the train set than in the test set) and in middle cerebral arteries territory (~10% more cases in the test set than in the train set) can be observed.

Technical Validation

The presented medical imaging data was derived from the picture archiving and communication system of the corresponding institutions and therefore fully complies with the legal standards and quality controls for the acquisition of medical imaging in Germany, the European Union and Switzerland, as well as the industrial standards of the scanner vendors. Segmentation masks were prepared and annotated at voxel-level by a human-machine hybrid algorithm with hierarchical manual checks and corrections first by specifically trained medical students and later a neuroradiology resident. Afterwards, the masks were reviewed, corrected, and finally approved by an expert neuroradiologist, ensuring a high quality standard.

We aimed to create a dataset which is representative of real world stroke cases. Therefore, only cases with massive motion artefacts, leading to images that are unusable in diagnostic clinical practice, were excluded from the dataset. We chose not to exclude cases based on other potential quality issues like signal loss or spatial distortions, relying on the imaging standards that are established in clinical practice at the participating centers. Similarly, no special regard was given to cases if they were acquired at 1.5 T as opposed to 3 T and the raters were blinded to mode of image acquisition.

Code availability

In order to facilitate future users of this dataset to get familiarized with the images, we have released the following ISLES 2022 Github repository: <https://github.com/ezequieldlrosa/isles22>. The repository contains scripts to read the images, visualize them, and to quantify the algorithmic results performance with the same metrics used in the challenge to rank participants. Besides, we have also released the image container that is used during the challenge in order to evaluate the participants' algorithm submissions (https://github.com/ezequieldlrosa/isles22_docker_evaluation).

Received: 4 August 2022; Accepted: 28 November 2022;

Published online: 10 December 2022

References

- Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2095–2128 (2012).
- Feigin, V. L. *et al.* Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* **383**, 245–254 (2014).
- Berkhemer, O. A. *et al.* A randomized trial of intraarterial treatment for acute ischemic stroke. *N. Engl. J. Med.* **372**, 11–20 (2015).
- Goyal, M. *et al.* Randomized assessment of rapid endovascular treatment of ischemic stroke. *N. Engl. J. Med.* **372**, 1019–1030 (2015).
- Jovin, T. G. *et al.* Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N. Engl. J. Med.* **372**, 2296–2306 (2015).
- Saver, J. L. *et al.* Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N. Engl. J. Med.* **372**, 2285–2295 (2015).
- Albers, G. W. *et al.* Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N. Engl. J. Med.* **378**, 708–718 (2018).
- Campbell, B. C. *et al.* Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N. Engl. J. Med.* **372**, 1009–1018 (2015).
- Ma, H. *et al.* Thrombolysis Guided by Perfusion Imaging up to 9 Hours after Onset of Stroke. *N. Engl. J. Med.* **380**, 1795–1803 (2019).
- Hjort, N. *et al.* Magnetic resonance imaging criteria for thrombolysis in acute cerebral infarct. *Stroke* **36**, 388–397 (2005).
- Thomalla, G. *et al.* DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4.5 h of symptom onset (PRE-FLAIR): a multicentre observational study. *Lancet Neurol.* **10**, 978–986 (2011).
- Ringleb P., *et al.* Akuttherapie des ischämischen Schlaganfalls, S2e-Leitlinie. in *Leitlinien für Diagnostik und Therapie in der Neurologie* (Deutsche Gesellschaft für Neurologie (Hrsg.), 2021).
- Rava, R. A. *et al.* Assessment of a Bayesian Vitrea CT Perfusion Analysis to Predict Final Infarct and Penumbra Volumes in Patients with Acute Ischemic Stroke: A Comparison with RAPID. *AJNR Am. J. Neuroradiol.* **41**, 206–212 (2020).

14. Xiong, Y. *et al.* Comparison of Automated CT Perfusion Softwares in Evaluation of Acute Ischemic Stroke. *J. Stroke Cerebrovasc. Dis.* **28**, 104392 (2019).
15. Mokin, M. *et al.* Predictive Value of RAPID Assessed Perfusion Thresholds on Final Infarct Volume in SWIFT PRIME (Solitaire With the Intention for Thrombectomy as Primary Endovascular Treatment). *Stroke* **48**, 932–938 (2017).
16. Rava, R. A. *et al.* Assessment of computed tomography perfusion software in predicting spatial location and volume of infarct in acute ischemic stroke patients: a comparison of Sphere, Vitrea, and RAPID. *J. Neurointerv. Surg.* **13**, 130–135 (2021).
17. Clerigues, A. *et al.* Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput. Biol. Med.* **115**, 103487 (2019).
18. de la Rosa, E., Sima, D.M., Kirschke, J.S., Menze, B. & Robben, D. Detecting CTP Truncation Artifacts in Acute Stroke Imaging from the Arterial Input and the Vascular Output Functions. *medRxiv*, 2022.2006.2016.22276371 (2022).
19. de la Rosa, E., Sima, D. M., Menze, B., Kirschke, J. S. & Robben, D. AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning. *Med. Image Anal.* **74**, 102211 (2021).
20. Ezequiel de la Rosa, D.R., Diana M. S, J S. Kirschke, B M. Differentiable Deconvolution for Improved Stroke Perfusion Analysis. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (2020).
21. Hakim, A. *et al.* Predicting Infarct Core From Computed Tomography Perfusion in Acute Ischemia With Machine Learning: Lessons From the ISLES Challenge. *Stroke* **52**, 2328–2337 (2021).
22. Robben, D. *et al.* Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Med. Image Anal.* **59**, 101589 (2020).
23. Freyschlag, C. F. *et al.* The Volume of Ischemic Brain Predicts Poor Outcome in Patients with Surgically Treated Malignant Stroke. *World Neurosurg.* **123**, e515–e519 (2019).
24. Meng, X. & Ji, J. Infarct volume and outcome of cerebral ischaemia, a systematic review and meta-analysis. *Int. J. Clin. Pract.* **75**, e14773 (2021).
25. Menze, B. H. *et al.* The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2015).
26. Zecavati, N. *et al.* The utility of infarct volume measurement in pediatric ischemic stroke. *J. Child. Neurol.* **29**, 811–817 (2014).
27. Maier, O. *et al.* ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med. Image Anal.* **35**, 250–269 (2017).
28. Ezequiel de la Rosa, U.H., *et al.* B.M.R.B. Ischemic Stroke Lesion Segmentation Challenge 2022: Acute, sub-acute and chronic stroke infarct segmentation. in *25th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2022)* (Zenodo, 2022).
29. Liew, S. L. *et al.* A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. *Sci. Data* **9**, 320 (2022).
30. Silversmith, W. cc3d: Connected components on multilabel 3D & 2D images. *Zenodo* (2021).
31. Allen, L.M., Hasso, A.N., Handwerker, J. & Farid, H. Sequence-specific MR imaging findings that are useful in dating ischemic stroke. *Radiographics* **32**, 1285–1297; discussion 1297–1289 (2012).
32. Lansberg, M. G. *et al.* Evolution of apparent diffusion coefficient, diffusion-weighted, and T2-weighted signal intensity of acute stroke. *AJNR Am. J. Neuroradiol.* **22**, 637–644 (2001).
33. Warach, S., Chien, D., Li, W., Ronthal, M. & Edelman, R. R. Fast magnetic resonance diffusion-weighted imaging of acute human stroke. *Neurology* **42**, 1717–1723 (1992).
34. Warach, S., Gaa, J., Siewert, B., Wielopolski, P. & Edelman, R. R. Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging. *Ann. Neurol.* **37**, 231–241 (1995).
35. Lutsep, H. L. *et al.* Clinical utility of diffusion-weighted magnetic resonance imaging in the assessment of ischemic stroke. *Ann. Neurol.* **41**, 574–580 (1997).
36. Schlaug, G., Siewert, B., Benfield, A., Edelman, R. R. & Warach, S. Time course of the apparent diffusion coefficient (ADC) abnormality in human stroke. *Neurology* **49**, 113–119 (1997).
37. Nagesh, V. *et al.* Time course of ADCw changes in ischemic stroke: beyond the human eye! *Stroke* **29**, 1778–1782 (1998).
38. Schwamm, L. H. *et al.* Time course of lesion development in patients with acute stroke: serial diffusion- and hemodynamic-weighted magnetic resonance imaging. *Stroke* **29**, 2268–2276 (1998).
39. Yang, Q. *et al.* Serial study of apparent diffusion coefficient and anisotropy in patients with acute stroke. *Stroke* **30**, 2382–2390 (1999).
40. Beaulieu, C. *et al.* Longitudinal magnetic resonance imaging study of perfusion and diffusion in stroke: evolution of lesion volume and correlation with clinical outcome. *Ann. Neurol.* **46**, 568–578 (1999).
41. Berndt, M. T. *et al.* Basal Ganglia versus Peripheral Infarcts: Predictive Value of Early Fiber Alterations. *AJNR Am. J. Neuroradiol.* **42**, 264–270 (2021).
42. Kaesmacher, J. *et al.* Early Thrombectomy Protects the Internal Capsule in Patients With Proximal Middle Cerebral Artery Occlusion. *Stroke* **52**, 1570–1579 (2021).
43. Schonfeld, M. H. *et al.* Effect of Balloon Guide Catheter Utilization on the Incidence of Sub-angiographic Peripheral Emboli on High-Resolution DWI After Thrombectomy: A Prospective Observational Study. *Front. Neurol.* **11**, 386 (2020).
44. Schonfeld, M. H. *et al.* Sub-angiographic peripheral emboli in high resolution DWI after endovascular recanalization. *J. Neurol.* **267**, 1401–1406 (2020).
45. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
46. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. arXiv:1606.06650 (2016).
47. Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
48. Fedorov, A. *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **30**, 1323–1341 (2012).
49. Isensee, F. *et al.* Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Mapp.* **40**, 4952–4964 (2019).
50. Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**, 196–205 (2010).
51. Shamonin, D. P. *et al.* Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front. Neuroinform* **7**, 50 (2013).
52. Schirmer, M. D. *et al.* Spatial Signature of White Matter Hyperintensities in Stroke Patients. *Front. Neurol.* **10**, 208 (2019).
53. Ourselin, S., Stefanescu, R. & Pennec, X. Robust Registration of Multi-modal Images: Towards Real-Time Clinical Applications. 140–147 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2002).
54. Hernandez Petzsche, M. R. *et al.* Probabilistic stroke lesion map of the ISLES'22 dataset. *Zenodo*. <https://doi.org/10.5281/zenodo.7335305> (2022).
55. Hernandez Petzsche, M. R. *et al.* ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Zenodo*. <https://doi.org/10.5281/zenodo.7153326> (2022).

Acknowledgements

MRHP is supported by TUM KKF Clinician Scientist Program. EDLR, MIM and IE are supported by the Translational Brain Imaging Training Network under the EU Marie Skłodowska-Curie program (Grant ID: 765148). BM, BW, and FK are supported through the SFB 824, subproject B12, by DFG through TUM International Graduate School of Science and Engineering, GSC 81, and by the Institute for Advanced Study, funded by the German Excellence Initiative. BM, BW, IE and FK acknowledge funding from DCoMEX (Grant agreement ID: 956201). BM is supported by Helmut Horten Foundation.

Author contributions

M.R.H.P. manuscript preparation, data acquisition, data review, segmentation corrections, image rating. E.D.L.R. code development, project design, data review, manuscript preparation. U.H., R.W., G.B., L.M. data acquisition, manuscript preparation, data review. B.M., W.V., M.R., M.I.M., S.L.L., F.K., I.E., D.R., A.H. data acquisition, code development, manuscript preparation, data review. T.F., T.Z., J.B., T.A.B. image rating, lesion segmentation. C.Z., T.B.B., M.B., B.I. data acquisition, manuscript preparation, data review. B.W. code development, segmentation correction, project design, data review, manuscript preparation. J.S.K. code development, data acquisition, segmentation correction, project design, data review, manuscript preparation.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

EDLR, DR and MIM are employed by icometrix. Independently of this work, TBB reports consultancy for MicroVention, Balt and Acandis and received speaker honoraria from Philips and phenox outside of the submitted work. CZ disclosed no relevant relationships regarding activities related to the present article. He has served on scientific advisory boards for Philips and Bayer Schering; serves as co-editor on the Advisory Board of Clinical Neuroradiology; has received speaker honoraria from Bayer-Schering and Philips; the institution has received research support and investigator fees for clinical studies from Biogen Idec, Quintiles, MSD Sharp & Dome, Boehringer Ingelheim, Inventive Health Clinical UK Ltd., Advance Cor, Brainsgate, Pfizer, Bayer-Schering, Novartis, Roche, Servier, Penumbra, WCT GmbH, Syngis, SSS International Clinical Research, PPD Germany GmbH, Worldwide Clinical Trials Ltd., Phenox, Covidien, Actelion, Medivation, Medtronic, Harrison Clinical Research, Concentric, Pharmtrace, Reverse Medical Corp., Premier Research Germany Ltd., Surpass Medical Ltd., GlaxoSmithKline, AXON Neuroscience, Bristol-Myers Squibb, Genentech, Acandis, EISAI, NeuroRx, Italfarmaco, Bioclinica, MIAC and IXICO. BW has received speaker honoraria from Philips. JSK is cofounder of Bonescreen GmbH, not related to this work. All other Authors report no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R.H.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022