Supplementary Materials and Methods: StrokeAnalyst's detailed methodology

S-2.1. Initial digital slice extraction from scanned brain images

As described in the section of the main manuscript "*Concept and methodology overview of "StrokeAnalyst", construction of the mouse TTC-atlas*", every brain is originally presented through a digital image of 2D coronal slices stored together (one next to each other) in one image I_a (Suppl. Figure 1a). In order to automatically identify the boundaries of each region that includes a TTC-slice within the I_a and extract each TTC-slice as single image, we first binarized the image I_a by a thresholding operation to roughly distinguish the TTC-slice tissue (foreground pixels) from background. 2D brain slices were identified by a function similar to the "analyze particles" function of ImageJ where groups of connected foreground pixels that exceeded a minimum brain area (large enough to represent a slice) were selected. An axis aligned bounding-box containing all pixels describing a 2D brain slice is calculated and used to extract (crop) the image region corresponding to each 2D slice (Suppl. Figure 1b).

The whole process was performed by a separate, in-house developed, MatLab-based, tool to "cut" and extract single TTC-slices named "BrainSlicer". (provided as a as link for downloading here: https://drive.google.com/drive/folders/1MrexC-oDJiXCLQr64I8hNj0jSalFVgX5?usp=sharing). Other commonly used software, such as ImageJ, can also produce the same results, nevertheless considerably more

laboriously. Eventually, each of these cropped images for each brain are stored in a separate file and used for the subsequent automated lesion analysis, as described in sections *"Stroke induction using the filament Middle Cerebral Artery Occlusion model (fMCAo)*" to "*Tissue collection, brain sectioning, slice TTC-staining and scanning from naive and stroked animals*".

S-2.2. Image preprocessing and background segmentation

Preprocessing of extracted TTC-slices (section S-2.1) begins with background extraction (Suppl. Figure 1c to 1d) that is fundamental for subsequent analysis of infarcts and hemisphere volumes. Let's consider the intensities of a TTC slice for each pixel $x \in R^2$ as a 2D image *I(x)*, and also let the set of segmentation labels be *background* (BG), *healthy* (HL) brain tissue (grey matter appearing in a red colors, white matter appearing in white colors) and *lesion* (LS) (infarcted brain tissue that appears in bright near-white colors). Since LS class shows a similar intensity profile with certain normal anatomical areas (HL class), such as the corpus callosum and all white matter tracts, the segmentation process has to distinguish between healthy and lesioned tissue. This is performed through multiple steps. Initially we divide the input image in superpixels to encapsulate neighborhood information[1]. K-means classification with three centers is then applied to assign a class label to every superpixel[2-4]. A Markov Random Field (MRF) is then superimposed and used to solve an energy minimization problem that favors maximally homogenous regions. This leads to reduction of small parts that were misclassified in the previous step[5]. If M_{BG} denotes the background mask, the complementary of it ($\overline{M_{BG}}$) is then used to extract the brain (healthy and infarcted) tissue., i.e. $I_{Brain}(x) = I(x) \circ \overline{M_{BG}}(x)$, where "$o$" is

the Hadamard product. Optionally, we can semi-manually improve the initial background segmentation (Suppl. Figure 1c), if necessary, using an active contour segmentation technique (ACS), as described in detail elsewhere[6-8], because some of the pixels close to the brain boundary may be misclassified as tissue instead of BG. Eventually, this process produces segmented (from their background) TTC-slices (I_s, s for segmented) as shown in Suppl. Figure 1d.

At last, spatial normalization of all segmented slices is performed, which includes image centering based on the center of mass and rotation around the center of mass, in order to remove differences in placement of the tissue samples in the scanner, and bring all samples in the same reference space. For global alignment we perform Principal Components Analysis[9] on the binary brain tissue mask and then rotate the corresponding image slice $I_{Brain}(x)$ according to the orientation indicated by the first principal component. This process aligns the TTC-slices as shown in Suppl. Figure 1e (final extracted image I_r, r for rotated) and prepares them for the subsequent stroke analysis (sections S-2.5 and S-2.6).

S-2.3. Image registration and TTC-atlas construction

For the needs of lesion analysis by SA, we developed a 2D TTC-atlas that encapsulates statistical information for the image intensity of healthy (pathology-free) brains at each anatomical location and is constructed as described next.

We inherited a coordinate system with anterior-superior-right orientation similar to Paxinos coordinates[10], where every slice is represented by a coronal plane at an index *j*, with *j = 0* lying at the plane of bregma (*j > 0* indicates a plane anterior to bregma and *j < 0* posterior to bregma). Eventually, every coronal plane (slice) of the

TTC-atlas will be also represented by its unique domain space $\Omega_A^J, j = 1, \dots, K$, where K is the number of slices of the TTC-atlas. Assuming that the image intensity for every pixel follows a Gaussian distribution, we can parametrize the distribution by the average and standard deviation values which together form the TTC-atlas. To calculate these values from given samples, the images have to be co-registered first so that each pixel corresponds to the same anatomical location across mouse brains. Therefore, for every coronal plane *j* we selected one of the available image slices as a template $I_A^j \in \Omega_A^j$ and registered the corresponding slices of the remaining mouse brains to it. For each of those moving images, denoted as $I_M^j \in \Omega_M^j$, the registration problem seeks to find a transformation $T^j: \Omega_M^j \to \Omega_A^j$ that spatially aligns the moving image to the reference image space. Omitting the index *j* for simplicity, we define the transformation *T(x)* as a composition of a linear (affine in our case) *($T^L(x)$)* and a non-linear (deformable) *($T^{NL}(x)$)* mapping, i.e. *$T(x) = T^{NL}(x) o\ T^L(x)$*. The linear component (global affine transformation[11]) accounts for translation, rotation and shearing differences due to the experimental setup and global variations in the mouse brain anatomies, whereas the non-linear component corrects local deformities. We estimate the linear transformation by maximizing the Mutual Information cost function[12, 13] using the Matlab image registration toolbox[14]. The non-linear deformable mapping is then estimated using the "Deformable Registration via Attribute Matching and Mutual-Saliency Weighting" (Dramms) software[15]. The same linear and non-linear registration algorithms (used for the healthy mouse brains) are also utilized to map images with lesion to the TTC-atlas space (section S-2.5).

Eventually, after the co-registration of all available brain slices at a given coronal plane, we compute the pixelwise average *(A_j)* and standard deviation *(SD_j)* maps for every index *j* across brains, and use them as the maximum-likelihood estimate parameters of each TTC-slice appearance. These two parametric maps form together the TTC statistical atlas for each coronal index (average and standard deviation images, *Aj* and *SDj*, Figure 1a). The methodology for the TTC-atlas calculation and construction is graphically illustrated in Figure 1a, whereas its utilization for lesion detection in new brains is shown in Figure 1b.

S-2.4. Definition of anatomical areas on the TTC-atlas

After the completion of the TTC-atlas construction we defined the anatomical regions at each TTC-atlas coronal plane. This is essential for the acquisition of neuroanatomical information of the lesioned areas. We map anatomical information from Allen Brain Atlas (ABA) [16] to the constructed TTC-atlas by transforming Allen's to TTC's atlas corresponding coronal slices. Due to significantly different tissue-processing modalities (paraformaldehyde fixed and dehydrated sections in Allen versus freshly-cut and non-dehydrated section in TTC), mapping could not be performed automatically but relied on a semi-automated registration process. Here, we manually annotated (landmarking) corresponding control points (>50) in the TTC-atlas and ABA over each slice and then utilized them to calculate a Local Weighted Mean transformation[17] that allowed to map ABA's anatomical masks to TTC-atlas space (Supplementary Figure 2).

S-2.5. Atlas-based abnormalities (lesion) detection

Brains from animals with stroke were processed as described in sections "*Tissue collection, brain sectioning, slice TTC-staining and scanning from naive and stroked animal*" of the main manuscript. Automated lesion analysis of stroke lesions on TTC-slices begins with manual indexing of each obtained TTC coronal slice (TTC-slices I_r named from now on as subject slice *V)* to its bregma coordinates (anteroposterior axis). Each subject slice *V* is then registered -fully automatically- to the corresponding TTC-atlas slice based on the procedure described in Section S-2.3 (i.e. through the application of a global affine [11] and a nonlinear transformation). This methodology produces a linear transformation matrix and a corresponding deformation field for each separate slice of the brain that we denote as T_1 and T_2 respectively (Figure 1b); both will further be used in the process.

Lesion detection is automatically performed in multiple steps. First, abnormality regions are detected based on the spatial likelihood of the tissue being healthy given its voxel-wise intensity, $In(p)$. We express the likelihood of observing the intensity *In* at pixel *p* if the brain is healthy, through the Z-score, $Z(p)$, as described by the equation:

$$Z(p) = \frac{In(p) - A(p)}{SD(p)},$$

where $A(p)$ and $SD(p)$ are the mean and standard deviation values at pixel *p* of the TTC-atlas. Thresholding of this abnormality score provides an initial lesion segmentation mask, $M_{LS}^{o}(p)$. The threshold value is determined empirically on brain slices with stroke. This mask includes many artifacts, mainly at the boundaries of the slice, due to the local staining procedure. These false positives have image intensities almost identical to that of the infarct. Thereby, for robust stroke lesion segmentation, further processing is required to remove all these contaminant pixels

(false-positives). This is performed by taking into account that in our experimental procedure only unilateral stroke lesions are considered, thus the identification of the affected hemisphere may allow to topologically constrain the lesion extent, thus reduce the search space. We consider as affected hemisphere the one with significantly more *abnormal tissue*, as quantified by the mask $M_{LS}^{o}(p)$.

Proper identification the two brain hemispheres is necessary for 1) identification of the affected hemisphere and proper further lesion detection (as mentioned above) and 2) for left-right hemisphere volume measurements (for edema calculation, section S-2.7). For hemispheres' segmentation the midline of each brain slice has to be extracted. For that, we seek and identify the axis of brain symmetry utilizing the Allen's anatomical masks. Midline calculation on the original Allen masks is a relatively simple process as ABA is a "rotation free" atlas constructed in perfect symmetry. The hemisphere masks are mapped from the Allen space to the TTC-atlas space. That said, we mask out the previous abnormality map retaining only the part in the affected hemisphere *($Z_H(p)$)* and introduce its values as feature characteristics for the next and final step of the lesion detection process, as presented next.

S-2.6. Lesion segmentation based on machine learning (ML)

During the atlas-based abnormality detection, "abnormality points" can depict either an actual lesion or local staining artifacts (corresponding to false positive points). For example, typical abnormality points with high *Z*-scores are often observed at the boundaries of each slice as an artifact of staining. We formulate a binary classification problem (see also section S-2.5) in which the candidate lesions (identified by *Z*-score thresholding) are classified into actual lesion (true) or non-

lesion tissue (false) and solve it using a random forest (RF) classifier[18]. We selected the RF classifier because it has shown to produce robust and accurate solutions in many application domains[19].

Three types of features are combined and introduced to RF:

- 3 intensity features from each of the 3 RGB color channels (thus 9 features in total), i.e. the original pixel intensity, and the values obtained after moving average filtering within a 3x3 kernel and after Gaussian filtering with *sigma = 2* for the kernel; the respective kernel parameters were experimentally optimized,
- two features expressing lesion probability in respect to healthy mouse brain population, i.e. the original probability value $Z_H(x)$ and the value after moving average filtering and
- 8 features expressing lesion probability in respect to brain symmetry. We distinguish these last 8 features into two groups, one based on the Z-score hemisphere difference and one based on the hemisphere difference calculated per color channel (RGB). For both groups (pure RGB values or Z-score values) we initially compute the hemisphere difference and then we extract two features for each color channel. As for the Z-score difference image, we follow the same procedure. These features are 1) the pure pixel intensity of the multiple difference images and the pixel value produced after a moving average filtering operation.

The combined 19 features are used to train an RF classifier by minimizing the expected misclassification cost among lesion and healthy tissue classes. The returned class label for every pixel *p* (arranged in 2D) eventually forms the ML-based lesion

segmentation mask (L_{RF} mask, Figure 1b). Finally, remaining false positives are further reduced by applying the *opening* morphological operation, i.e. erosion followed by dilation[20], to remove isolated small pixel regions and to produce our final lesion segmentation for each brain slice.

The final lesions masks (L_F) at each coronal index are eventually used for (automated) stroke volumetry and neuroanatomical region identification. To calculate the affected area, we need to map each L_F back into subject's V native space. This procedure is performed by inverting the transformations obtained during spatial normalization, T_1^{-1} and T_2^{-1} (as described in section S-2.5.). Let the affected area in slice j be $S_j, j = 1, \dots, K$, where K is the number of slices, and the number of non-zero pixels in the lesion mask L_F to be $R_j,\ j = 1, .., K$. We calculate the lesion area as $S_j = R_j \cdot pixel\_size$, where $pixel\_size = 0.021 \cdot 0.021 = 0.000441\ mm^2$ is the set value for images that are scanned at 1200dpi (standard resolution). Hence, by multiplying with the slice thickness, and summing over all affected slices, we calculate the lesion volume for each mouse brain.

S-2.7. Infarct volume and edema calculations

We calculated the volume of left/right hemisphere and infarct in each slice (respective areas × actual thickness of slice) and then the total (brain) volumes of left/right hemispheres and infarct, as they are derived from manual and automated processes. Infarct volume is calculated as previously described[2, 4], as a percentage of the left (non-ischemic) hemisphere (%Vinf), corrected for edema or atrophy using the formula:

$$\%Vinf = \frac{(V\ right\ hemisphere + V\ infarct - V\ left\ hemisphere)}{V\ right\ hemisphere} * 100\ .$$

The % hemispheric edema is calculated using the formula:

$$\%HE = \frac{V\ left\ hemisphere - V\ right\ hemisphere}{V\ right\ hemisphere} * 100\ .$$

References

1. Verma N, Cowperthwaite MC, Markey MK. Superpixels in brain MR image analysis. *Annu Int Conf IEEE Eng Med Biol Soc* 2013; 2013**:** 1077-80.

2. Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa LDF *et al.* Clustering algorithms: A comparative approach. *PLoS One* 2019; 14(1)**:** e0210236.

3. Steinley D. K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 2006; 59(Pt 1)**:** 1-34.

4. Juang L-H, Wu M-N. MRI brain lesion image detection based on color-converted K-means clustering segmentation. *Measurement* 2010; 43(7)**:** 941-949.

5. Paragios N, Ferrante E, Glocker B, Komodakis N, Parisot S, Zacharaki EI. (Hyper)-graphical models in biomedical image analysis. *Med Image Anal* 2016; 33**:** 102-106.

6. Lee H, Codella NCF, Cham MD, Weinsaft JW, Wang Y. Automatic Left Ventricle Segmentation Using Iterative Thresholding and an Active Contour Model With Adaptation on Short-Axis Cardiac MRI. *IEEE Transactions on Biomedical Engineering* 2010; 57(4)**:** 905-913.

7. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process* 2001; 10(2)**:** 266-77.

8. Caselles V, Kimmel R, Sapiro G. Geodesic Active Contours. *International Journal of Computer Vision* 1997; 22(1)**:** 61-79.

9. Rehman HZU, Lee S. Automatic Image Alignment Using Principal Component Analysis. *IEEE Access* 2018; 6**:** 72063-72072.

10. Fraklin KBJ, Paxinos G. *The Mouse Brain in Stereotaxic Coordinates*, 3rd Edition edn Academic Press, Elsevier, 2007.

11. Mandal PK, Mahajan R, Dinov ID. Structural brain atlases: design, rationale, and applications in normal and pathological cohorts. *J Alzheimers Dis* 2012; 31 Suppl 3**:** S169-88.

12. Woo J, Stone M, Prince JL. Multimodal registration via mutual information incorporating geometric and spatial context. *IEEE Trans Image Process* 2015; 24(2)**:** 757-69.

13. Mattes D, Haynor D, Vesselle H, Lewellyn T, Eubank W. *Nonrigid multimodality image registration*, vol. 4322. SPIE, 2001.

14. The MathWorks I. MATLAB 2018a. In. Natick, Massachusetts, United States., 2018.

15. Ou Y, Sotiras A, Paragios N, Davatzikos C. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Med Image Anal* 2011; 15(4)**:** 622-39.

16. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 2007; 445(7124)**:** 168-76.

17. Image enhancement method using piecewise linear transforms. *2004 12th European Signal Processing Conference*; 6-10 Sept. 2004, 2004.

18. Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M *et al.* Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? *Brief Bioinform* 2013; 14(3)**:** 315-26.

19. Hermann M, Schunke AC, Schultz T, Klein R. Accurate Interactive Visualization of Large Deformations and Variability in Biomedical Image Ensembles. *IEEE Trans Vis Comput Graph* 2016; 22(1)**:** 708-17.

20. Ong KH, Ramachandram D, Mandava R, Shuaib IL. Automatic white matter lesion segmentation using an adaptive outlier detection method. *Magn Reson Imaging* 2012; 30(6)**:** 807-23.