# Ischemic Stroke Lesion Segmentation

www.isles-challenge.org

## Proceedings

5th October 2015

Munich, Germany

UNIVERSITÄT ZU LÜBECK

TUM
Technische Universität München

$u^b$
UNIVERSITÄT BERN

MICCAI15
MUNICH

## Preface

Stroke is the second most frequent cause of death and a major cause of disability in industrial countries. In patients who survive, stroke is generally associated with high socioeconomic costs due to persistent disability. Its most frequent manifestation is the ischemic stroke, whose diagnosis often involves the acquisition of brain magnetic resonance (MR) scans to assess the stroke lesion's presence, location, extent, evolution and other factors. An automated method to locate, segment and quantify the lesion area would support clinicians and researchers alike, rendering their findings more robust and reproducible.

New methods for stroke segmentation are regularly proposed. But, more often than desirable, it is difficult to compare their fitness, as the reported results are obtained on private datasets. Challenges aim to overcome these shortcomings by providing (1) a public dataset that reflects the diversity of the problem and (2) a platform for a fair and direct comparison of methods with suitable evaluation measures. Thus, the scientific progress is promoted.

With ISLES, we provide such a challenge covering ischemic stroke lesion segmentation in multi-spectral MRI data. The task is backed by a well established clinical and research motivation and a large number of already existing methods. Each team may participate in either one or both of two sub-tasks:

**SISS**    Automatic segmentation of ischemic stroke lesion volumes from multi-spectral MRI sequences acquired in the sub-acute stroke development stage.

**SPES**    Automatic segmentation of acute ischemic stroke lesion volumes from multi-spectral MRI sequences for stroke outcome prediction.

The participants downloaded a set of training cases with associated expert segmentations of the stroke lesions to train and evaluate their approach, then submitted a short paper describing their method. After reviewing by the organizers, a total of 17 articles were accepted and compiled into this volume. At the day of the challenge, each teams' results as obtained on an independent test set of cases will be revealed and a ranking of methods established.

For the final ranking and more information, visit WWW.ISLES-CHALLENGE.ORG.

*Oskar Maier, Universität zu Lübeck*

*Mauricio Reyes, University of Bern*

*Björn Menze, TU Munich*

*August 2015*

**Organizers**

Oskar Maier, Universität zu Lübeck, Germany

Mauricio Reyes, University of Bern, Switzerland

Björn Menze, TU Munich, Germany


**Sponsoring Institutions**

Institute of Medical Informatics, Universität zu Lübeck, Germany

Institute for Surgical Technology & Biomechanics, University of Bern, Switzerland

Computer Science, TU Munich, Germany

# ISLES (SISS) challenge 2015: segmentation of stroke lesions using spatial normalization, Random Forest classification and contextual clustering

Hanna-Leena Halme[1,2], Antti Korvenoja[1], and Eero Salli[1]

[1] HUS Medical Imaging Center, Radiology, University of Helsinki and Helsinki
University Hospital, P.O. Box 340, FI-00029 HUS, Finland
[2] Department of Neuroscience and Biomedical Engineering NBE, Aalto University
School of Science, P.O. Box 12200, FI-00076 Aalto, Finland

## 1 Statistical template

In order to enable voxel-wise comparisons across subjects, all T1, T2, Flair and DWI volumes, as well as the volumes containing manual segmentations, were warped to a common template. The template was generated by the Advanced Normalization Tools (ANTs) software version 2.1.0rc3 [1] using *buildtemplateparallel* script, greedy SyN transformation model, cross-correlation similarity metric, $30 \times 90 \times 20$ interations and the T1 images of the training dataset. All T1 images were deformed to this template with ANTs software using Affine transformation model for rigid registration and SyN transformation model for warping. The detailed parameters are given in Appendix. This defined a transformation which was applied to T2, Flair, DWI and manual segmentation volumes. All MR images were masked with a volume containing only in-brain voxels to minimize the effect of background to the automatic segmentation of ischemic stroke lesion volumes.

Images in the common template space representing mean and standard deviation (std) of voxel intensities over subjects were calculated voxel-by-voxel, separately for T1, T2, Flair and DWI images. We call these images *statistical* templates from now on. Note that the lesion voxels were not included in the calculation of average and std. Because most subjects in the training dataset had lesions on the left hemisphere, there were fewer voxels contributing to the mean and std on the left; as a result, the left hemisphere appeared slightly distorted on the template images. In order to compensate for this left-right bias, the mean and std images were additionally averaged over left and right hemispheres. Furthermore, the images were smoothed with a 3D Gaussian kernel (FWHM 3 mm) to decrease the effect of registration inaccuracies.

## 2 Random Forest classification algorithm

The initial segmentation was predicted with an ensemble learning method. A set of features was derived from the training data and fed to a Random Forest

[2] classification algorithm implemented in Scikit-learn version 0.16.dev [3]. The Random Forest algorithm combines classification results from a number of decision trees. Several trees are constructed and fitted to the data during training phase, using a random subset of features to train each tree. The final classification is the mode of the classes obtained from all individual trees. Random Forest classification greatly reduces overfitting, which is a common problem for simple decision tree classifiers [2]. In this study, the performance of the classifier was tested with leave-one-out cross-validation, in which one subject from the training dataset was used for testing and the rest for training the classifier, and the procedure was repeated for all subjects. The classifier returned both the binary classifications and probabilities (in range 0-1) that a voxel belonged to the lesion area.

## 3 Classifier training

16 features were extracted from the MR images for subsequent classification. Z-score normalized voxel intensities of T1, T2, Flair and DWI images constituted features 1-4. Features 5-8 represented the Z-score deviations from global average images, calculated separately for each sequence by subtracting the global mean and dividing with the global std. The purpose of these features was to find regions showing large deviations from the normal brain, which likely indicates presence of a lesion. Features 9-12 were obtained by smoothing the original images with a 3D Gaussian kernel (FWHM 3 mm), thus including information from the local neighborhood of each voxel. Smoothing was expected to improve classification since it may reduce the effect of registration inaccuracies. Features 13-16 represented local asymmetry, obtained by comparing voxel intensities on one hemisphere to the corresponding voxel intensities on the other hemisphere. The motivation for calculating local asymmetry was the fact that lesions rarely occur symmetrically on both hemispheres. The asymmetry measure was calculated simply by subtracting the original smoothed image from the left-right -mirrored smoothed image.

In order to decrease computational time and avoid classifier overfitting, we only collected the aforementioned features from a randomly selected subset of voxels. The maximum number of lesion voxels sampled from each subject was set to 300, and in cases where the lesion was smaller than that, all lesion voxels were sampled. The ratio of lesion and non-lesion voxels per subject was kept constant, such that twice as many voxels were sampled from non-lesion area as from lesion area; thus, the maximum number of non-lesion voxels per subject was 600.

For the Random Forest classifier, the training set was resampled to train a total of 300 decision trees. 4 features were used to obtain the best split at each individual tree. The quality of each split was described by Gini impurity [2]. The trees were grown unlimitedly, i.e. until each leaf contained only samples of a single class. All parameters and default values used by Scikit-learn's RandomForest classifier are listed in Appendix.

# 4  Contextual clustering

The segmentation results obtained with the Random Forest method were further improved with contextual clustering (CC). The clustering method was based on a Markov random fields (MRF) prior and iterated conditional modes (ICM, [4]) algorithm, which were previously used for analysis of functional magnetic resonance imaging (fMRI) data [5]. The basic assumption in contextual clustering is that neighboring voxels tend to belong to the same class. Furthermore, it is assumed that the intensity distribution of background voxels (in present case, non-lesion voxels) is standard normal, but the distribution of lesion voxels is unknown.

The CC algorithm modified for this study consisted of the following steps:

1. Fit a gamma distribution to all nonzero voxels of the probability map given by the random forest classifier. Fitting is done using MATLABs (R2015a) function *fitdist* with default parameters.
2. Transform the probability map values to standard normal distribution by calculating the inverse normal distribution function (MATLAB function *norminv*) from the cumulative distribution function (MATLAB function *cdf*) of gamma distribution. This gives image $N$.
3. Define the parameter $T$ for contextual clustering [5] using the fitted gamma distribution:
   $T = -norminv(cdf(gamma, D))$, where $gamma$ is the fitted gamma distribution, and $D$ some threshold. In this study we used empirically chosen $D = 0.6$, which gave reasonable results with training data.
4. Run the CC algorithm [5] using the image $N$, neighborhood weight coefficient $\beta = T^2/6$ and a threshold $D = 0.6$. The voxels will be reclassified to 0 or 1, corresponding to non-lesion and lesion, respectively.
5. Repeat steps 1-4 with only the voxels classified as non-lesion in the first run of CC.

Finally, all automatically segmented images were transformed back to each subject's native space using inverse transformation and nearest neighbor interpolation. After transformation it was possible to compare the automatic segmentations with the manual lesion segmentations. The classification accuracy was evaluated with the script provided at ISLES web page (http://www.isles-challenge.org), including measurements for Dice coefficient, average symmetric surface distance (ASSD), Hausdorff distance, precision and recall.

# 5  Testing phase

The test data will be spatially normalized to the common template using the parameters listed in Appendix, but without the lesion images (–x option). After that, the trained Random Forest classifier and contextual clustering will be applied to the data.

# References

1. Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage, 54(3), 2033-2044.
2. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
3. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825-2830.
4. Besag, J. (1986). On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B (Methodological), 259-302.
5. Salli, E., Aronen, H. J., Savolainen, S., Korvenoja, A., & Visa, A. (2001). Contextual clustering for analysis of functional MRI data. Medical Imaging, IEEE Transactions on, 20(5), 403-414.

# Appendix

### Generation of templates
The common template was done in two phases. First, the initial template was formed:
buildtemplateparallel.sh -d 3 -m 1x0x0 -n 0 -r 1 -t GR -s CC -o [initial template image] -c 0 -j 1 [T1 images]

After that, the final template was built using the initial template:
buildtemplateparallel.sh -d 3 -m 30x90x20 -n 0 -r 0 -t GR -s CC -o [template image] -z [initial template image] -c 0 [T1 images]

Warping of T1 images to common template was done using *antsRegistration* tool and the following parameters:
–metric MI[template image, T1 image,1,32] –transform affine[0.25] –convergence 10000x10000x10000x10000x10000 –shrink factors 5x4x3x2x1 –smoothing-sigmas 4x3x2x1x0 –metric CC[template image, T1 image,1,5] –transform SyN[0.25,3.0,0.0] –convergence 50x35x15 –shrink factors 3x2x1 –smoothing-sigmas 2x1x0 -use-histogram-matching 1 –x [lesion image]

### Parameters for Random Forest classifier
Scikit-learn's function sklearn.ensemble.RandomForestClassifier was used with the following parameters:
n_estimators=300, criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=4, max_leaf_nodes=None, bootstrap=True, oob_score=False, n_jobs=1, random_state=None, verbose=0, warm_start=False, class_weight=None