

Head CT Data Release

Description of the images shared:

The data consists of 35 different subject's non-contrast CT scan, with 2 subjects having 2 scans but every other subject having one scan. One subject additionally has 2 scans, but from the same time point, but with different convolution kernels. Each scan has at least one reader's manual segmentation of the image to delineate the mask of the brain areas (including cerebrospinal fluid (CSF)).

The data are from the Minimally-Invasive Surgery plus rt-PA for Intracerebral hemorrhage Evacuation (MISTIE) Phases II and III clinical trials the Clot Lysis: Evaluating Accelerated Resolution of Intraventricular Hemorrhage (CLEAR) Phase III clinical trial. Each subject's scans and DICOM (Digital Imaging and Communications in Medicine) header information are contained in a tarball (`.tar.xz`), with the name of the tarball being the `id` of the subject (`01.tar.gz` is for `id 01`).

Each Tarball

Image Data

Each tarball contains a series of NIfTI (Neuroimaging Informatics Technology Initiative) files. These files can be read using common tools such as ITK-SNAP (<http://www.itksnap.org/pmwiki/pmwiki.php>), Mango (<http://ric.uthscsa.edu/mango/>), or Papaya (<http://ric.uthscsa.edu/mango/papaya.html>). Each image has manual segmentation from 2 different readers which is a binary image delineating the brain and an automated estimate of the brain by the method proposed in the paper below.

The NIfTI images are as follows:

1. `BRAIN_1_Anonymized.nii.gz` - CT image that has been de-identified/anonymized where the face and the ears have been removed from the image. These should have a range of -1024 to 3071 and be in Hounsfield Units (HU).
2. `BRAIN_1_Anonymized_Mask.nii.gz` - a binary image denoting the brain areas using the method described in Muschelli et al. (2015). This method is also implemented in the function `CT_Skull_Strip` function in the `ichseg` R package (<https://github.com/muschellij2/ichseg>).
3. `Manual_Mask_1_Reader_1.nii.gz` - binary brain mask from Reader 1
4. `Manual_Mask_1_Reader_2.nii.gz` - binary brain mask from Reader 2

Reader 1 and reader 2 are consistent across subjects (e.g. reader 1 is always reader 1).

Header information

In the tarballs, there are also `CSV` (Comma-separated) files, such as:

- `BRAIN_1_header.csv`
- `Manual_Mask_1_Reader_1_header.csv`
- `Manual_Mask_1_Reader_2_header.csv`

which contain information from the DICOM header. Mostly the only files of interest are those with

BRAIN/BONE in the name as the Manual mask images are derived from those images. These are included for completeness.

Specific subsets of the DICOM data/tags are given with the data: 0008-0070-MANUFACTURER 0018-0050-SLICETHICKNESS 0018-0060-KVP 0018-1120-GANTRYDETECTORTILT 0018-1151-XRAYTUBECURRENT 0018-1152-EXPOSURE 0018-1160-FILTERTYPE 0018-1170-GENERATORPOWER 0018-1210-CONVOLUTIONKERNEL 0018-5100-PATIENTPOSITION 0020-0032-IMAGEPOSITIONPATIENT 0020-1041-SLICELOCATION 0028-0030-PIXELSPACING 0008-0070-Manufacturer 0018-0050-SliceThickness 0020-0032-ImagePositionPatient 0020-1041-SliceLocation 0028-0030-PixelSpacing

Different files in some tarballs

Subject 13 has a BONE scan as well as a BRAIN scan. The BONE scan is a non-contrast CT done with a different convolution kernel, which gives different properties of the scan. The manual segmentations are in the same space as both scans (i.e. no registration was done).

Subjects 06 and 16 have both a BRAIN_1 and BRAIN_2 scan. These were images taken at 2 different time points. If the manual reader performed segmentation on the scan, then a file such as Manual_Mask_2_Reader_1.nii.gz would indicate the index for the scan (e.g. Manual_Mask_2 corresponds to BRAIN_2).

Demographics

Demographic information is id column – which maps back to our original de-identified ID, which has no information about patient other than site and enrollment number, but the mapping is not provided or released. Additional information such as age in years (range 40 – 78) rounded to nearest year, sex, race (as measured in the trial), Hispanic indicator, site ID (again maps back to the original data but no mapping is provided) and diagnosis (dx). the diagnosis was made from the reading center, indicating intracerebral hemorrhage (ICH) or ICH with intraventricular hemorrhage (IVH), indicated as ICH with IVH in the dx column.

The site ID is to allow researchers to see if there is a small site effect (but not likely possible due to how few subjects).

The de-identification mechanism

A mask was determined outlined on a template of a mask and the ears of the image. The face and ear masking of the data was done by registering a template image to the individual images separately. The method is similar to pydeface (<https://github.com/poldracklab/pydeface>) and was implemented in the fs1r (<https://github.com/muschellij2/fs1r>) package. The template data was adapted from pydeface and the template used for registration was https://github.com/muschellij2/pydeface/raw/master/pydeface/data/mean_reg2mean.nii.gz and the template mask used was https://github.com/muschellij2/pydeface/raw/master/pydeface/data/facemask_no_ears.nii.gz.

References

Muschelli, John, et al. "Validated automatic brain extraction of head CT images." *Neuroimage* 114 (2015): 379-385. <https://doi.org/10.1016/j.neuroimage.2015.03.074>