

## MRI brain imagery processing software in data analysis

Alexander Bernstein<sup>1</sup>, Akzhigitov Renat<sup>1,2</sup>, Ekaterina Kondrateva<sup>1</sup>, Svetlana Sushchinskaya<sup>1</sup>, Irina Samotaeva<sup>1,2,3</sup>, Vladislav Gaskin<sup>1,2</sup>

<sup>1</sup>Skolkovo Institute of Science and Technology, Moscow, Russia.

<sup>2</sup>Moscow Research and Clinical Center for Neuropsychiatry, Moscow, Russia.

<sup>3</sup>Laboratory of applied physiology of higher nervous activity,  
Institute of higher nervous activity and neurophysiology, RAS, Moscow, Russia.

a.bernstein@skoltech.ru,  
ekaterina.kondrateva@skoltech.ru,  
svetlana.sushchinskaya@skolkovotech.ru

**Abstract.** Modern brain scanning techniques are one of the most consistent sources of medical information which included in most diagnostic protocols. These data possess a great potential for the machine learning (ML) analysis and clinical applications of the results. MRI imaging is prevalent in brain analysis, allowing acquisition of structural (3D image) and functional imaging (timeseries of 3D images). Relative to the MRI data volumes these research are considered to the big data and massive analysis, connected with numerous methods of signal processing, advanced ML approaches and feature extraction. In current paper we update the surveys of most software used for the feature generation, processing and analysis of MRI brain imagery. Systematizing the toolboxes, we describe common scanner characteristics, image modalities, corresponding features and their informational context, as well as common machine learning classification problems in working with the data. And as a result, we propose evaluation and comparison of conventional software tools with their detailed description.

**Keywords:** MRI Software Packages, MRI toolbox, Machine learning, Preprocessing, Classification, Neuroscience.

### 1 Introduction

Much progress has been made over the last years in computational neuroscience. The storage, visualization, processing and analysis of the medical imagery data are not feasible without computer-based techniques.

There is a constantly growing number of machine learning studies based on neuroimaging data. These research aimed to develop diagnostic and clinical prognostic

prediction tools, facilitate an automatic segmentation or contribute in understanding the mechanisms of diseases by revealing new biomarkers and patterns.

A problem of machine learning classification in MRI imagery is usually related to multistage pipelines with handcrafted feature generation in different software packages, feature selection and analysis by methods of machine learning. Thus, the ML results are directly connected to the scanner characteristics, image modality and content, processing software and generated features quality.

Currently, there are various software packages, each targeted on different analysis and providing numerous features for users. Since there is no single package that can provide all of possible informative features, it is helpful to know the output of each software package, as well as the internal software characteristics [1].

In this paper, we provide the general information concerning different MRI modalities with related features, open databases and comparison of several software packages for feature generation, which are frequently used for the classification.

## **2 MRI data analysis**

### **2.1 Imaging in neuroscience**

The visualization in neuroscience could be achieved with usage of various scanning methods, covering Magnetic-Resonance Imaging (MRI), X-ray Computed Tomography (CT), Molecular imaging: Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT), Diffuse Optical Tomography (DOT), High-Resolution EEG (HR-EEG), Magnetoencephalography (MEG), Ultrasound imaging and their combinations [2].

Among of these methods the Magnetic-Resonance Imaging and Computed tomography are the most widely spread in clinical practice. The CT is beneficial in patients with implantable medical devices, such as cardiac pacemakers, vascular clips and cochlear implants as for these patients MRI scanning is contraindicated; in disturbed or very heavy patients, because the imaging can be performed much more rapidly, comparable to MRI.

In current paper we are focused on the MRI imagery because in most cases it is preferred in brain imaging due to the greater range of available tissue contrast, detailed anatomy and high sensitivity to abnormalities [3]. There is the superiority of MRI analysis in children and patients undergoing several imaging examinations explained by its potential harmlessness.

### **2.2 Medical Image Data and Datasets in terms of machine learning**

The huge potential of machine learning approaches was proven in image-based diagnosis, risks evaluation and clinical prognosis [4], [5].

Despite this, currently in clinical practice each patients scan is observed individually by a radiologist. For this approach the massive data analysis is not required: all imagery can be processed on one local machine (for visualization, morphometry estimation, etc.).

In contrast, the machine learning application to the brain imagery data is a massive data analysis. Then storing, managing and processing the 3D images with millions of voxels, or 4D sequences containing the timeseries for each voxel multiplied on the number of recordings and different acquisition modalities, became critical.

Using the machine learning one can process the original data in 3D arrays, but the most popular approach is to generate the vector of *features* consisted of hundred/thousands characteristics responsible for the specific brain property and proceed the data for the classification; most frequently these vectors are generated by the MRI brain analysis software.

It should be mentioned, that the size and quality of training dataset have the dominant influence on a ML model performance. Currently neuroscience research doesn't have large enough datasets (millions of samples) and the data collecting and sharing are critical in order to learn as much as we can about the neurological conditions [6].

There are still open questions concerning the collection, annotation, discovery, and sharing or reuse of medical imaging data.

### 2.3 Major MRI imaging properties

Existed MRI scanners provide various *spatial characteristics, magnetic field strengths* as well as different *imaging protocols*.

*Spatial characteristics*: major scanner features are represented by the spatial in-plane resolution, voxel size and a slice thickness. These characteristics mostly depend on the scanner inductive detection coil parameters [7].

*Magnetic Field*: Nowadays, there are scanners with from 1.5T to 7T magnetic field strength in clinical practice [8]. The higher magnetic strength field exposes higher signal-to-noise and contrast-to-noise ratio, higher resolution, enhanced contrast images which is beneficial in diverse pathologies, like multiple sclerosis, tumors, degenerative and age-related changes. Thus, additional information can be gained compared to imaging with lower field strength, in cause of better depiction of small anatomical details [9].

However, imaging at high field as 7T is tackled with several limitations as: the fields inhomogeneity, a higher specific absorption rate and additional contraindications for patient scanning and most importantly: higher field tomography is not widely spread [10].

In lower strength fields (1.5T vs 3T) the results of scanning may also differ. It was shown that the quality of 3T brain scan can be more feasible, but in most cases the informativity is comparable to 1.5T scan [11].

*Imaging protocols*: there are plenty of acquisition protocols and the number is evolving in parallel to hardware development and modernization of signal processing methods. The overall set could be divided to two broad categories: anatomical (structural) and functional scans.

In addition to the variety of protocols in different scanners, there is a possibility to manually adjust the protocol, for instance changing the echo time (TE), repetition time (TR) or switching the acquisition mode from axial to sagittal.

The characteristics listed above will directly affect the image quality and content; influence of these parameters on the processing results are still under the investigation. Thus, conducting the research, it is crucial to think of the dataset homogeneity and consider the level of the related uncertainty in case if the dataset contains imagery from different sources.

In this regard, there is a need to stick the research reporting and data managing protocols for the MRI/fMRI study and completely describe the used procedure and apparatus.

## 2.4 Structural MRI features

The structural imaging is aimed to provide the imagery for the static brain structure, enhancing desired patterns with usage of different acquisition modalities.

The anatomical scan types vary by acquisition time (AT), the time to echo (TE) and repetition time (TR) weighting: shorter TR relates to so-called T1 (spin-lattice) relaxation and affects contrast on T1-weighted images, longer TE relates to T2 (spin-spin) decay and affects contrast on T2-weighted images; fluid-attenuated inversion recovery (FLAIR) - by very long TR and TE times, and proton density weighting - by long TR and short TE, showing the difference in brain tissue physical characteristics. Structural scan can also be diffusional, or perfusion weighted (DWI, PWI), susceptibility weighted pointing the distortion of local magnetic field (SWI), specialized or enhanced (magnetization-prepared rapid gradient-echo (MP-RAGE), etc.).

The canonical ones for the structural imaging are: T1-weighted structural image (T1), T2-weighted fluid-attenuated inversion recovery structural image (T2 FLAIR), diffusion-weighted imaging (DWI), susceptibility-weighted imaging (SWI). Their usage in clinical practice is prescribed in most diagnostic protocols and it was proven that these 4 sequences are sufficient for the evaluation of patients with a new neurological complaint [12].

Features: sMRI images contain the anatomical (morphometric) features explaining the brain structural characteristics: gray matter surface area, cortical volumes thicknesses, sulcal depth, curvatures in sulcal regions, and sulcal walls. These data can be vectorized to the one-dimensional explanation of patients' cortex and subcortical nuclei (based on T1 or T2-weighted structural imagery).

Specifically patterns of calcification and hemorrhage could be enhanced by in the SWI imagery and then recognized more correctly by machine learning methods of segmentation.

White matter tracts (tractography) contains an information of so called structural connectivity reflecting which regions of interest (ROI) are structurally connected between each other. These data could be represented in graphs as well as in vectorized and then processed to the ML (DWI, DTI imagery)

The functional unit commonly used in brain imagery analysis is the region of interest: a subset of voxels identified for a specific purpose. These ROIs are arranged in different atlases parceling the brain tissue according to its known or hypostatized functionality [13].

## 2.5 Functional MRI features

The functional MRI investigation used for the study of the brain function based on measure of the brain neural activation reflected in blood oxygen-level dependent signal (BOLD signal). The fMRI study is not represented in most counties clinical practice, but in some – included in presurgical protocols [14]; otherwise conducted mostly for the research purposes.

For the functional imagery there are two types being frequently represented: resting-state functional MRI time series (rfMR) and task functional MRI time series data (tMRI). They could be observed in tandem or only rsMRI separately.

Most fMRI scans are gradient-echo spin-echo planar imaging (GE-EPI or SE-EPI) and therefore are particularly sensitive to artifacts in ROI typically affected by susceptibility effects.

*Features:* The fMRI series represented a series of compressed (in several folds lower resolution) 3D images. Popular functional connectivity analysis is described as a level of signal correlation between desired ROIs, highly correlated regions in BOLD are considered as functionally connected or involved in one neurocognitive process.

Thus, there is a possibility to observe 1D time series for desired ROIs, to conduct an analysis using the functional connectivity symmetrical matrixes or perform a graph-based analysis on these data.

Functional connectivity (fMRI) under the masking of structural connectivity (DWI or DTI) or hd-EEG and MEG will result in multimodal effective connectivity analysis [15].

The atlases from structural imaging could be applied to functional one considering its lower dimensionality.

## 3 Machine learning in MRI neuroimaging

As it was stated above, the machine learning is in a great request in many fields of study for biomedical tasks. The classical machine learning problems are classification and regression. As an input machine learning classifier get sample data and then, as an output give the probability of belonging to a particular class [16]–[22].

Modern machine learning techniques are quite successful for picture-based tasks. Recent Deep Convolutional Neural Networks were created under the biological inspiration and provide efficient data analysis [23]. Many diagnosis and disease prognosis objectives were explored and yielded new finding with modern data analysis techniques; for instance, in case of Alzheimer’s Disease [24]–[28].

There are two main methods for picture-based tasks considering MRI and fMRI: to work with initial picture or to generate features from the images. If we do not generate features there is a problem with the interpretability of results.

Machine learning in neuroscience is quite complicated task because of high dimensionality of the data and relatively small dataset. For diagnostic aims it is easier to work with interpreted features. In this case we consider the vector representation of the brain with interpretable parameters *features*, discussed above. For instance, for the MRI such features are volumes, thicknesses, intensities, curvatures and surface areas of different

brain regions and application of classical machine learning algorithms such as support vector machines (SVC), Random Forest (RFC), k nearest neighbors (kNC) or logistic regression (LR) used for classification.

The classical approach for machine learning task for identifying the patterns and trends is to obtain or generate features and then train classifiers on generated data. For this purpose, usually, external software toolboxes are used. As nowadays iPython is the most powerful instrument for data analysis it is more convenient use toolboxes implemented in iPython. One of the most popular project integrated to iPython is Nipy. It contains not only Analysis Pipeline Management for the most popular Neuroimaging tools, but also Nilearn - fast and easy way of statistical learning on neuroimaging data.

Since neuroimages have complex structure, there is a need for neuroimage based problems special ML approaches development. Often before analyzing neuroimages dimensionality reduction is needed, before application of classification approaches. Such methods of manifold learning as kernel principal component analysis(kPCA) are represented in most of the basic languages environments libraries (iPython, Matlab, etc.).

Another complication of neuroimaging tasks is overfitting. This happened because of small datasets for training and the sophisticated data structure. The overfitting might be controlled by cross-validation process - technique for estimating the performance of machine learning algorithms. The variety of cross-validation techniques are implemented in iPython (Sklearn lib) as well as in Matlab.

The most modern way of exploring initial data are deep learning algorithms. For the MRI or fMRI classification 3D or 4D Convolutional Neural Networks could be used. These algorithms could be used for large number of data. They provide powerful pipeline for the local feature generating of brain patterns and special techniques as batch normalization should be implemented to avoid overfitting.

## 4 MRI processing software

There are hundreds of available toolboxes/add-ints/tools for MRI data managing (BIDS), converting (Nibabel), visualization (3D Slicer), masking (MRICron), atasing (AAL), preprocessing (MRIQC, Mindboggle, etc.), preprocessing and analysis (SPM, FSL etc.), specialized analysis (Dype, CONN, LST, VLSM, etc.) predefined Pipeline resources (data, executable modules, workflows and services).

The features, observed in the proposed comparison, were chosen the way to give a newcomer to fMRI data analysis an overview of existed software packages potentials from the very general information to more detailed.

The list of packages with diverse usage and their basic information is given in the Table 1.

The medical significance of the toolbox is indirectly confirmed by the frequency of use of relevant articles; the reference papers for most popular open-source toolboxes and the number of their citations are represented in the Table 2.

The Table 3 contains the detailed information for the most popular software.

In the Table 4 we collated the description of the software/services providing pipelines for the several toolboxes.

The information in the tables is represented according to the following characteristics:

- URL address: the working and up-to-date URL supporting the software package;
- Use: the targeted purpose of the software; could be divided into;
- Availability: the information whether the package is an open source or commercially distributed;
- Referencing paper/Announcement date: the release date information, citing paper;
- Number of references: retrieved from Google Scholar search up to the 15 February 2018;
- Support: the supporting organization, foundation or university;
- Platform/operating system: software and hardware required for the installation;
- Preprocessing: consisted of the raw data obtained from the MRI scanner corrections, including artifact detection, spatial realignment, movement correction, centering, normalization, and smoothing in one pipeline that could be manually digested;
- Data input: the software availability to work with the 2D, 3D, and 4D images;
- Atlasing: working with the region of interest (ROI) input, as amount of default preinstalled atlases, the possibility of manual atlas editing or working with the custom one;
- Statistical analysis: if the package has internal module for the statistical analysis;
- Extensions: prescribed add-on extensions increasing the software functionality diversity and specificity.

**Table 1.** The List of the software for brain MRI data managing, processing and analysis.  
*V- visualization, P- preprocessing, A- analysis (including group analysis);*  
*O- open source, C- commercially distributed.*

| Package       | Reference URL   | Use    | Access |
|---------------|---|--------|--------|
| AFNI          | <a href="http://afni.nimh.nih.gov/afni/">http://afni.nimh.nih.gov/afni/</a>     | V,P,A  | O      |
| Analyze       | <a href="http://www.analyzedirect.com/">http://www.analyzedirect.com/</a>       | A      | C      |
| ANTS          | <a href="http://www.picsl.upenn.edu/ANTS/">http://www.picsl.upenn.edu/ANTS/</a> | V,P, A | O      |
| Brain Voyager | <a href="http://www.brainvoyager.com/">http://www.brainvoyager.com/</a>         | V, A   | C      |
| DCMTK         | <a href="http://www.dcmtdk.org/">http://www.dcmtdk.org/</a>                     | P      | O      |

MRI brain imagery processing software in data analysis 10

|                           |   |        |   |
|---------------------------|---|--------|---|
| E-Prime                   | <a href="http://www.pstnet.com/eprime.cfm">http://www.pstnet.com/eprime.cfm</a>   | A      | C |
| fMRI Power                | <a href="http://www.fmripower.org/">http://www.fmripower.org/</a>   | A      | O |
| Freesurfer                | <a href="http://surfer.nmr.mgh.harvard.edu/">http://surfer.nmr.mgh.harvard.edu/</a>   | V,P,A  | O |
| FSL                       | <a href="http://www.fmrib.ox.ac.uk/fsl/">http://www.fmrib.ox.ac.uk/fsl/</a>   | V,P,A  | O |
| ITK                       | <a href="http://www.itk.org/">http://www.itk.org/</a>   | A      | O |
| MRI Studio                | <a href="https://www.mristudio.org/">https://www.mristudio.org/</a>   | V      | O |
| MRICron                   | <a href="http://www.cabiatl.com/mricro/mricro/mricro.html">http://www.cabiatl.com/mricro/mricro/mricro.html</a>   | P      | O |
| Presentation              | <a href="http://www.neurobs.com/">http://www.neurobs.com/</a>   | P      | C |
| SPM                       | <a href="http://www.fil.ion.ucl.ac.uk/spm/">http://www.fil.ion.ucl.ac.uk/spm/</a>   | V,P, A | C |
| 3D Slicer                 | <a href="https://www.slicer.org/">https://www.slicer.org/</a>   | V,P    | C |
| SpinalCord Toolbox        | <a href="https://sourceforge.net/projects/spinalcordtoolbox/">https://sourceforge.net/projects/spinalcordtoolbox/</a>   | A      | O |
| Nilearn                   | <a href="http://nilearn.github.io/">http://nilearn.github.io/</a>   | A      | O |
| BioimageSuite             | <a href="http://bioimagesuite.yale.edu/">http://bioimagesuite.yale.edu/</a>   | A      | O |
| BrainMagix                | <a href="https://www.imagilys.com/brainmagix-clinical-neuroimaging-software-suite">https://www.imagilys.com/brainmagix-clinical-neuroimaging-software-suite</a> | P      | C |
| FMRLAB                    | <a href="https://scn.ucsd.edu/fmrlab/">https://scn.ucsd.edu/fmrlab/</a>   | A      | O |
| GraphVar                  | <a href="https://www.nitrc.org/projects/graphvar/">https://www.nitrc.org/projects/graphvar/</a>   | A      | O |
| ISAS                      | <a href="https://nonlinearestimation.bitbucket.io/">https://nonlinearestimation.bitbucket.io/</a>   | P      | O |
| AIR                       | <a href="http://www.cabi.gatech.edu/mricro/mricro/">http://www.cabi.gatech.edu/mricro/mricro/</a>   | P      | O |
| MRtrix3                   | <a href="http://www.mrtrix.org/">http://www.mrtrix.org/</a>   | A      | O |
| NIAK                      | <a href="http://niak.simexp-lab.org/">http://niak.simexp-lab.org/</a>   | P      | O |
| Tortoise                  | <a href="https://www.nitrc.org/projects/tortoise/">https://www.nitrc.org/projects/tortoise/</a>   | P      | O |
| Humain Connectome Project | <a href="http://www.humanconnectomeproject.org/">http://www.humanconnectomeproject.org/</a>   | A      | O |

|                |   |   |   |
|----------------|---|---|---|
| Ciftify(HCP)   | <a href="https://github.com/edickie/ciftify">https://github.com/edickie/ciftify</a>         | P | O |
| GIFT           | <a href="http://mialab.mrn.org/software/gift/">http://mialab.mrn.org/software/gift/</a>     | A | O |
| BART           | <a href="https://mrirecon.github.io/bart/">https://mrirecon.github.io/bart/</a>             | P | O |
| CAT12          | <a href="http://www.neuro.uni-jena.de/cat/">http://www.neuro.uni-jena.de/cat/</a>           | P | O |
| CIVET          | <a href="https://mcin-cnim.ca/technology/civet/">https://mcin-cnim.ca/technology/civet/</a> | P | O |
| Neuro3D        | <a href="https://www.healthcare.siemens.com/">https://www.healthcare.siemens.com/</a>       | P | C |
| NordicNeuroLab | <a href="http://www.nordicneurolab.com/">http://www.nordicneurolab.com/</a>                 | P | O |

**Table 2.** Number of references to open-source software, obtained from the Google Scholar Search up to 15 February 2018.

| Toolbox       | Software Name (Reference to the original paper)  | # citations |
|---------------|--|-------------|
| SPM           | Statistical parametric maps in functional imaging: A general linear approach(K.J. Friston, 1995)   | 8924        |
| AFNI          | Software for analysis and Visualization of Functional Magnetic Resonance Neuroimages.(Robert W. Cox, 1996)   | 6366        |
| FSL           | Advance in functional and structural MR image analysis and implementation as FSL. (Stephen M. Smith, 2004)   | 6987        |
| FreeSurfer    | FreeSurfer: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain.(B. Fischel, 2002)   | 3961        |
| AIR           | Automated Image Registration: I. General Methods and Intrasubject, Intramodality Validation.(Roger P. Woods, 1998)   | 1893        |
| MRICRO        | Stereotaxi display of brain lesions. (C. Rorden, 2000)   | 1701        |
| CONN          | A functional connectivity toolbox for correlated and anticorrelated brain networks. (Whitfield-Gabrieli, S., and Nieto-Castanon, A. 2012)  | 650         |
| Brain Voyager | Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis (R. Geobel, 2006) | 778         |
| Ciftify (HCP) | The minimal preprocessing pipelines for the Human Connectome Project. (Glasser MF, 2013)   | 603         |
| VOXBO         | Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. (M Botvinick, 2005)   | 584         |
| BrainSuite    | BrainSuite: An automated cortical surface identification tool  | 564         |
| PLS           | Partial least squares analysis of neuroimaging data applications and advances.(A. R. McIntosh,2004)  | 561         |
| fMRLAB        | Single-Trial Variability in Event-Related BOLD Signals (J. R. Duann, 2002)   | 221         |

|                           |   |     |
|---------------------------|---|-----|
| BrainVISA                 | Software platform for visualization and analysis of multi-modality brain data(Y. Cointeps, 2001)  | 127 |
| NIH Toolbox               | Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. Lancet Neurol. 2010;9(2):138-139. | 134 |
| Nilearn                   | Machine learning for neuroimaging with scikit-learn (A.Abraham, 2014)   | 120 |
| Humain Connectome Project | Informatics and data mining tools and strategies for the Human Connectome Project. (Marcus, D.S.,2011)  | 113 |
| DPTools                   | Whole Brain quantitative CBF and CBV measurement using MRI plus tracking: comparison of methodologies.(Anne M. Smith, 2000)   | 85  |

**Table 3.** Detailed information for most popular and conventional software.

| Features/Toolbox           | SPM12   | Freesurfer   | FSL  | Nilearn (Nipy)  |
|----------------------------|---|--|--|---|
| Base language/<br>software | Matlab  | Stand alone  | Sland alone  | Python  |
| Input data                 | Structural MRI<br>Functional MRI<br>Diffusion MRI<br>(add ins)  | Structural MRI<br>Functional MRI<br>Diffusion MRI  | Structural MRI<br>Functional MRI<br>Diffusion MRI  | Structural MRI<br>Functional MRI<br>Diffusion MRI<br>(add ins)  |
| Main Functions             | Slice timing correction<br>Spatial normalisation<br>Segmentation<br>Canonical Variates Analysis<br>Regional Bayesian model comparison<br>Mixed-effects models<br>Bayesian model selection for group studies | Skull stripping<br>B1 bias field correction,<br>gray-white matter segmentation<br>Reconstruction of cortical surface models<br>Labeling of regions on the cortical surface<br>Nonlinear registration of the cortical surface of an individual with a stereotaxic atlas | model-based analysis<br>quantitative resting perfusion<br>analysis from perfusion<br>ASL FMRI<br>structural analysis<br>brain extraction<br>tissue-type segmentation<br>segmentation of subcortical structures | Functional Connectivity<br>Automatic Dataset Fetching<br>Decoding<br>Multivariate decompositions<br>Image processing and resampling utilities<br>Loading and Processing files easily<br>Data Masking<br>Utilities<br>Operating on regions<br>Mass-univariate analysis |

|                                       |   |   |  |                                |
|---------------------------------------|---|---|--|--------------------------------|
|                                       | Reparametrisation of the bilinear model         | Statistical analysis of group morphometry differences | voxelwise analysis of multi-subject grey-matter density longitudinal | Plotting brain data            |
| Operating system                      | Linux, Mac OS, Windows                          | Linux, Mac OS X Windows (via VirtualBox)              | Linux, Mac OS, Windows Virtual Machine                               | Linux, Mac OS, Windows         |
| Source code availability              | yes   | yes   | yes  | yes                            |
| Add ins                               | yes   | yes   | yes  | yes                            |
| Support                               | Founded   | NITRC   | NITRC  | NITRC                          |
| Statistical analysis                  | yes   | yes   | yes  | yes                            |
| Atlases supported                     | Desikan-Killiany Destrieux Yeo Volumetric atlas | Desikan-Killiany Destrieux Yeo                        | Desikan-Killiany Destrieux Yeo                                       | Desikan-Killiany Destrieux Yeo |
| Data analysis                         | Matlab tools                                    | No  | No   | Python tools                   |
| Google requests per 2017 #Science tag | 168   | 995   | 3820   | 77                             |

**Table 4.** List of software/interfaces providing the pipelines for several toolboxes.

| Source   | Usage    | Pipeline   | URL   |
|----------|----------|--|---|
| Nipype   | Software | FSL, ANTs, FreeSurfer, AFNI, SPM, Camino, MRtrix, Slicer, MNE etc      | <a href="http://nipype.readthedocs.io/en/latest/">http://nipype.readthedocs.io/en/latest/</a>           |
| fMRIPrep | Software | FSL, ANTs, FreeSurfer, AFNI.   | <a href="http://fmripiprep.readthedocs.io/en/latest/#">http://fmripiprep.readthedocs.io/en/latest/#</a> |
| NDMG     | Software | FSL, Dipy, Nibabel, Nilearn, Networkx, Numpy, Scipy, Scikit-Learn etc. | <a href="https://github.com/neurodata/ndmg">https://github.com/neurodata/ndmg</a>                       |

|           |                       |  |   |
|-----------|-----------------------|--|---|
| MRIQC     | Image quality metrics | FSL, ANTs, AFNI.                                     | <a href="http://mriqc.org">http://mriqc.org</a>                           |
| LONI      | Web-based interface   | Freesurfer, SPM, BrainSuit, Diffusional Toolkit etc. | <a href="http://pipeline.loni.usc.edu/">http://pipeline.loni.usc.edu/</a> |
| OpenNeuro | Web-based interface   | Freesurfer, SPM, Diffusional Toolkit etc.            | <a href="https://openneuro.org/">https://openneuro.org/</a>               |

#### 4.1 Software reproducibly testing and claims

For some of the software tool boxes it is well known that results are sensitive operating system and the program version installed [29]–[31]. Therefore, is crucial to check the release of the program, because the software failure in the step of feature generation will lead to discarding all the research results.

To get the reliable and reproducible data software users asked to:

- Keep updated with the latest version of the software;
- Conduct all the research on one machines, or within the machines cluster with similar architectures.

## 5 Conclusion

Machine learning approaches have a potential to fundamentally alter medical image analysis via development diagnostic and clinical prognostic prediction tools, but it requires the specific approach to the data processing.

In this paper we observed the peculiarities of work with the MRI imagery software, as the most widely spread method for interpretable feature generation for further ML analysis.

We provide a general review on the most available toolboxes for the magnetic resonance imagery analysis; common MRI data types, properties and modalities.

There is a growing number of software packages for MRI data analysis representing general or more specialized functionality. We did a comparison of existed software, updated the results of previously published surveys (see Table 1).

The worldwide research are mostly done by the free available methods, because it allows overall comparison, reproducing and direct knowledge transfer. In Table 2, we summarized the number of references for open-source software, which showed conventional suits for processing and analyzing human brain MRI data, including SPM, FreeSurfer, FSL, ANTs, AFNI etc.

The summary on the number of these software properties are represented in Table 3. Also, recently appeared resources for the data processing pipelines, such as OpenNeuro and NiPype, allowing usage and comparison of several mentioned before toolboxes on a cite, were listed in a Table 4.

We mentioned questions of MRI research reporting and dataset homogeneity, results verification and reproducibility; gave our perspective view on the features, obtained from the MRI data processing, their informative content and discussed common machine learning classification problems in working with the MRI data.

**Acknowledgments.** This work was supported by Skoltech "Biomedical Initiative Program".

## References

- [1] M. Behroozi and M. R. Daliri, "Software tools for the analysis of functional magnetic resonance imaging," *Basic Clin. Neurosci.*, vol. 3, no. 5, pp. 71–83, 2012.
- [2] H. Ammari, "An introduction to mathematics of emerging biomedical imaging," *Mathématiques et Applications*, vol. 62. Berlin: Springer, 2008.
- [3] J. S. Duncan, "Neuroimaging of the epilepsies," *Riv. di Neuroradiol.*, vol. 17, no. 3, pp. 446–463, 2004.
- [4] M. de Bruijne, "Machine learning approaches in medical image analysis: From detection to diagnosis," *Med. Image Anal.*, vol. 33, pp. 94–97, 2016.
- [5] M. D. Kohli, R. M. Summers, and J. R. Geis, "Medical Image Data and Datasets in the Era of Machine Learning—Whitepaper from the 2016 C-MIMI Meeting Dataset Session," *J. Digit. Imaging*, vol. 30, no. 4, pp. 392–399, 2017.
- [6] F. Alfaro-Almagro *et al.*, "Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank," *Neuroimage*, vol. 166, no. October 2017, pp. 400–424, 2018.
- [7] E. Kaza, U. Klose, and M. Lotze, "Comparison of a 32-channel with a 12-channel head coil: Are there relevant improvements for functional imaging?," *J. Magn. Reson. Imaging*, vol. 34, no. 1, pp. 173–183, 2011.
- [8] J. Neafie, "1.5 T," vol. 479, no. 7373, pp. 1–14, 2015.
- [9] H. C. Moon, H.-M. Baek, and Y. S. Park, "Comparison of 3 and 7 Tesla Magnetic Resonance Imaging of Obstructive Hydrocephalus Caused by Tectal Glioma.," *Brain tumor Res. Treat.*, vol. 4, no. 2, pp. 150–154, 2016.
- [10] A. G. Van Der Kolk, J. Hendrikse, J. J. M. Zwanenburg, F. Visser, and P. R. Luijten, "Clinical applications of 7 T MRI in the brain," *Eur. J. Radiol.*, vol. 82, no. 5, pp. 708–718, 2013.
- [11] R. Wood, K. Bassett, Foerster, C. Spry, and L. Tong, "1.5 Tesla Magnetic Resonance Imaging Scanners Compared With 3.0 Tesla Magnetic Resonance Imaging Scanners: Systematic Review of Clinical Effectiveness.," *CADTH Technol. Overv.*, vol. 2, no. 2, p. e2201, 2012.
- [12] W. A. Mehan *et al.*, "Optimal brain MRI protocol for new neurological complaint," *PLoS One*, vol. 9, no. 10, 2014.
- [13] M. Bilello *et al.*, "Population-based MRI atlases of spatial distribution are specific to patient and tumor characteristics in glioblastoma," *NeuroImage Clin.*, vol. 12, pp. 34–40, 2016.
- [14] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional

- networks,” *PLoS Comput. Biol.*, vol. 3, no. 2, pp. 0174–0183, 2007.
- [15] S. Klamer *et al.*, “Multimodal effective connectivity analysis reveals seizure focus and propagation in musicogenic epilepsy,” *Neuroimage*, vol. 113, pp. 70–77, 2015.
- [16] T. S. Kellermann *et al.*, “Prediction of specific depressive symptom clusters in youth with epilepsy: The NDDI-E-Y versus Neuro-QOL SF,” pp. 1–10, 2017.
- [17] B. Hosseinfard, M. H. Moradi, and R. Rostami, “Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal,” *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 339–345, 2013.
- [18] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, “Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG,” *Proc. 2008 IEEE Work. Mach. Learn. Signal Process. MLSP 2008*, pp. 244–249, 2008.
- [19] M. J. Patel, A. Khalaf, and H. J. Aizenstein, “Studying depression using imaging and machine learning methods,” *NeuroImage Clin.*, vol. 10, pp. 115–123, 2016.
- [20] T. Pekkala *et al.*, “Development of a late-life dementia prediction index with supervised machine learning in the population-based CAIDE study,” *J. Alzheimer’s Dis.*, vol. 55, no. 3, pp. 1055–1067, 2017.
- [21] M. Khondoker, R. Dobson, C. Skirrow, A. Simmons, and D. Stahl, “A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies,” *Stat. Methods Med. Res.*, vol. 25, no. 5, pp. 1804–1823, 2016.
- [22] J.-G. Lee *et al.*, “Deep Learning in Medical Imaging: General Overview,” *Korean J. Radiol.*, vol. 18, no. 4, p. 570, 2017.
- [23] P. Thodoroff, J. Pineau, and A. Lim, “Learning Robust Features using Deep Learning for Automatic Seizure Detection,” vol. 56, 2016.
- [24] S. Korolev, A. Safullin, M. Belyaev, and Y. Dodonova, “Residual and plain convolutional neural networks for 3D brain MRI classification,” *Proc. - Int. Symp. Biomed. Imaging*, pp. 835–838, 2017.
- [25] M. Park and W.-J. Moon, “Structural MR Imaging in the Diagnosis of Alzheimer’s Disease and Other Neurodegenerative Dementia: Current Imaging Approach and Future Perspectives,” *Korean J. Radiol.*, vol. 17, no. 6, p. 827, 2016.
- [26] A. D. Dekker, J. Fortea, R. Blesa, and P. P. De Deyn, “Cerebrospinal fluid biomarkers for Alzheimer’s disease in Down syndrome,” *Alzheimer’s Dement. Diagnosis, Assess. Dis. Monit.*, vol. 8, pp. 1–10, 2017.
- [27] D. Kim *et al.*, “Knowledge-driven binning approach for rare variant association analysis: application to neuroimaging biomarkers in Alzheimer’s disease,” *BMC Med. Inform. Decis. Mak.*, vol. 17, no. S1, p. 61, 2017.
- [28] E. Imabayashi *et al.*, “Comparison between brain CT and MRI for voxel-based morphometry of alzheimer’s disease,” *Brain Behav.*, vol. 3, no. 4, pp. 487–493, 2013.
- [29] T. Glatard *et al.*, “Reproducibility of neuroimaging analyses across operating systems,” *Front. Neuroinform.*, vol. 9, no. April, pp. 1–14, 2015.
- [30] C. Velasco-Annis, A. Akhondi-Asl, A. Stamm, and S. K. Warfield, “Reproducibility of Brain MRI Segmentation Algorithms: Empirical Comparison of Local MAP PSTAPLE, FreeSurfer, and FSL-FIRST,” *J. Neuroimaging*, 2017.
- [31] D. L. Tudorascu *et al.*, “Reproducibility and bias in healthy brain segmentation: Comparison of two popular neuroimaging platforms,” *Front. Neurosci.*, vol. 10, no.

Alexander Bernstein, Akzhigitov Renat, Ekaterina Kondrateva et. al 17

NOV, pp. 1–8, 2016.