Elsevier Editorial System(tm) for Computers in Biology and Medicine

Manuscript Draft

Manuscript Number: CBM-D-17-00197R2

Title: ARIANNA: a Research Environment for Neuroimaging Studies in Autism Spectrum Disorders

Article Type: Full Length Article

Keywords: Magnetic Resonance Imaging; Neuroimaging; Autism Spectrum Disorders; Children; Resource Sharing; Data Processing; World Wide Web Working Environment.

Corresponding Author: Dr. Alessandra Retico, PhD

Corresponding Author's Institution: Istituto Nazionale di Fisica Nucleare

First Author: Alessandra Retico, PhD

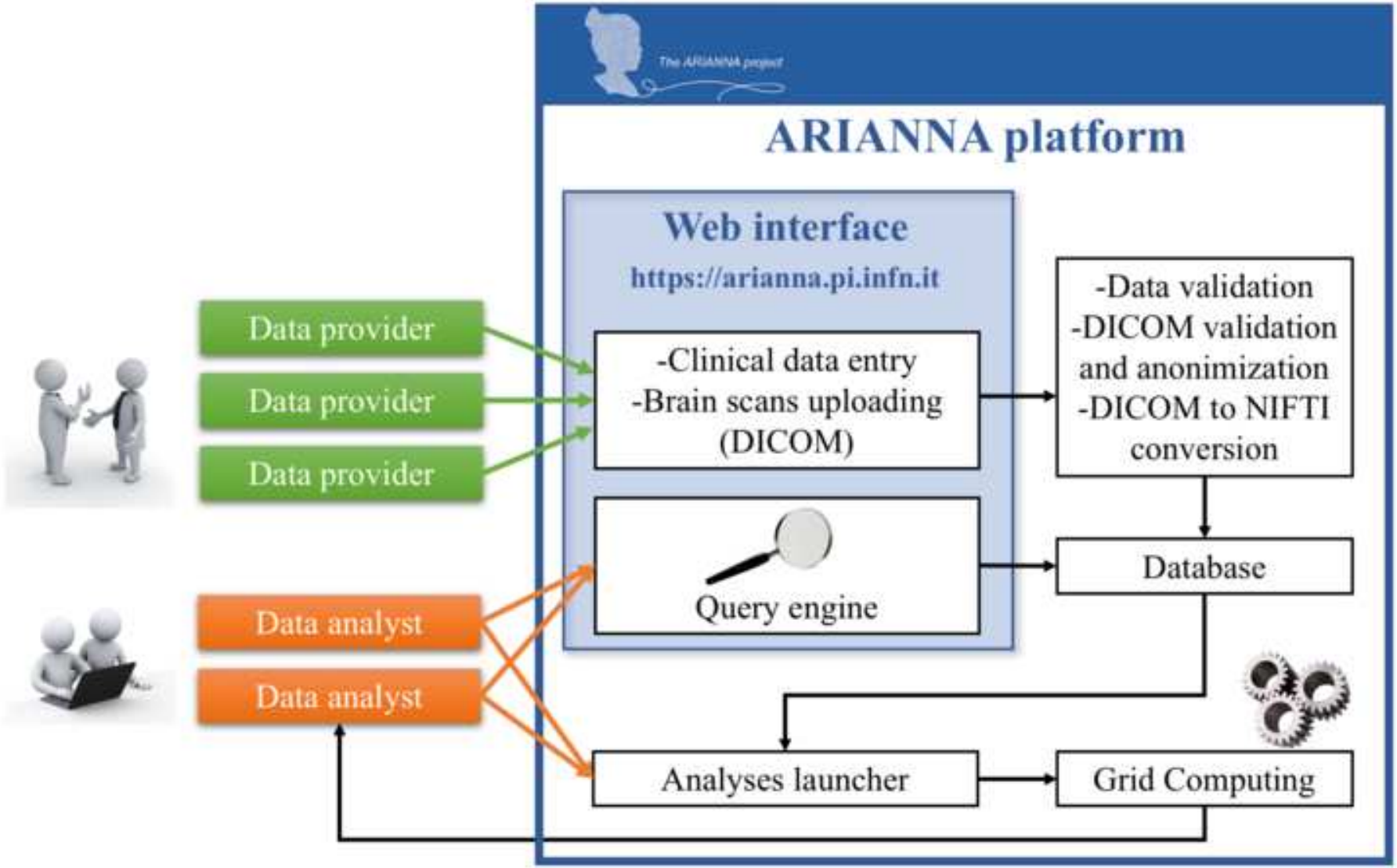
Order of Authors: Alessandra Retico, PhD; Silvia Arezzini; Paolo Bosco; Sara Calderoni; Alberto Ciampa; Simone Coscetti; Stefano Cuomo; Luca De Santis; Dario Fabiani; Maria Evelina Fantacci; Alessia Giuliano; Enrico Mazzoni; Pietro Mercatali; Giovanni Miscali; Massimiliano Pardini; Margherita Prosperi; Francesco Romano; Elena Tamburini; Michela Tosetti; Filippo Muratori

Abstract: The complexity and heterogeneity of Autism Spectrum Disorders (ASD) require the implementation of dedicated analysis techniques to obtain the maximum from the interrelationship among many variables that describe affected individuals, spanning from clinical phenotypic characterization and genetic profile to structural and functional brain images. The ARIANNA project has developed a collaborative interdisciplinary research environment that is easily accessible to the community of researchers working on ASD (https://arianna.pi.infn.it). The main goals of the project are: to analyze neuroimaging data acquired in multiple sites with multivariate approaches based on machine learning; to detect structural and functional brain characteristics that allow the distinguishing of individuals with ASD from control subjects; to identify neuroimaging-based criteria to stratify the population with ASD to support the future development of personalized treatments. Secure data handling and storage are guaranteed within the project, as well as the access to fast grid/cloud-based computational resources. This paper outlines the web-based architecture, the computing infrastructure and the collaborative analysis workflows at the basis of the ARIANNA interdisciplinary working environment. It also demonstrates the full functionality of the research platform. The availability of this innovative working environment for analyzing clinical and neuroimaging information of individuals with ASD is expected to support researchers in disentangling complex data thus facilitating their interpretation.

Conflict of Interest Statement

None Declared

Graphical Abstract (for review)



Highlights

+ARIANNA is a web-based research environment (arianna.pi.infn.it) aimed to:
+collect and analyze neuroimaging and clinical data acquired in multiple sites;
+detect structural and functional brain changes related to Autism Spectrum Disorders.
+ARIANNA guarantees secure data handling and storage.
+Fast grid/cloud-based computational resources are accessible.

ARIANNA: a Research Environment for Neuroimaging Studies in Autism Spectrum Disorders

Alessandra Retico[1*], Silvia Arezzini[1], Paolo Bosco[1], Sara Calderoni[2,3], Alberto Ciampa[1], Simone Coscetti[1], Stefano Cuomo[4], Luca De Santis[5], Dario Fabiani[1], Maria Evelina Fantacci[1,6], Alessia Giuliano[1], Enrico Mazzoni[1], Pietro Mercatali[4], Giovanni Miscali[5], Massimiliano Pardini[5], Margherita Prosperi[2], Francesco Romano[4], Elena Tamburini[7], Michela Tosetti[2], Filippo Muratori[2,3]

[1] National Institute for Nuclear Physics (INFN), Largo Bruno Pontecorvo 3, 56127 Pisa, Italy
[2] IRCCS Stella Maris Foundation, Viale del Tirreno 331, 56128 Pisa, Italy
[3] Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy
[4] Institute of Legal Information Theory and Techniques (ITTIG) of the National Research Council, Via de' Barucci 20, 50127 Florence, Italy
[5] NET7 S.r.l., via Marche 10, 56123 Pisa, Italy
[6] University of Pisa, Physics Department, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy
[7] I+ S.r.l., Piazza Puccini 26, 50144 Florence, Italy

* Corresponding author: Alessandra Retico, PhD; National Institute for Nuclear Physics (INFN), Largo Bruno Pontecorvo 3, 56127 Pisa, Italy; E-Mail: Alessandra.Retico@pi.infn.it; Phone: +39 050 2214459; Fax: +39 050 2214460.

Abstract

The complexity and heterogeneity of Autism Spectrum Disorders (ASD) require the implementation of dedicated analysis techniques to obtain the maximum from the interrelationship among many variables that describe affected individuals, spanning from clinical phenotypic characterization and genetic profile to structural and functional brain images. The ARIANNA project has developed a collaborative interdisciplinary research environment that is easily accessible to the community of researchers working on ASD (https://arianna.pi.infn.it). The main goals of the project are: to analyze neuroimaging data acquired in multiple sites with multivariate approaches based on machine learning; to detect structural and functional brain characteristics that allow the distinguishing of individuals with ASD from control subjects; to identify neuroimaging-based criteria to stratify the population with ASD to support the future development of personalized treatments. Secure data handling and storage are guaranteed within the project, as well as the access to fast grid/cloud-based computational resources. This paper outlines the web-based architecture, the computing infrastructure and the collaborative analysis workflows at the basis of the ARIANNA

interdisciplinary working environment. It also demonstrates the full functionality of the research platform. The availability of this innovative working environment for analyzing clinical and neuroimaging information of individuals with ASD is expected to support researchers in disentangling complex data thus facilitating their interpretation.

Keywords

Magnetic Resonance Imaging; Neuroimaging; Autism Spectrum Disorders; Children; Resource Sharing; Data Processing; World Wide Web Working Environment.

1-Introduction

Autism Spectrum Disorders (ASD) consist in a heterogeneous group of neurodevelopment disorders characterized by impaired social-communication function and the presence of restrictive and repetitive behaviors or interests, all of which significantly impact adaptive functioning [1]. Recent epidemiological studies have detected a prevalence of ASD in the population of the United States as high as 1 person in 68 [2], much higher than previous estimates, possibly reflecting general knowledge of ASD in the general public and in primary care professionals, diagnostic substitution, the increasing recognition of clinically milder cases, as well as the influence of environmental factors [3]. Despite it having been demonstrated that young children with ASD benefit from early intervention, when brain plasticity is maximal and environmental variables may have major effects on neurodevelopment [4][5], early diagnosis remains challenging for clinicians [6].

Common areas of impairment define the ASD condition, however ASD individuals show large phenotypic heterogeneity in severity of core symptoms, adaptive function, cognitive and language abilities, and medical/psychiatric comorbidities [7]. Moreover, a series of neurodevelopment disorders [e.g. language disorder, social (pragmatic) communication disorder, attention-deficit/hyperactivity disorder, stereotypic movement disorder] could show partially overlapping clinical manifestations and thus should be considered in the differential diagnosis of ASD patients.

To address this complex scenario, researchers have been intensively searching for ASD biomarkers. They are expected to complement or improve the current behavioral ASD diagnosis, to favor the early recognition of the pathological condition, and, possibly, to uncover the underling neurobiological causes of ASD [8].

Among the candidate biomarkers are the gene expression profile [9], the proteomic profile [10], the metabolomics profile [11], the head circumference trajectory [12][13], different measures of brain size, structure and function [14] and alterations in preferential looking tasks [15][16]. Despite none of them currently demonstrating sufficient sensitivity and specificity to be of clinical utility, all have served to achieve scientific advances in understanding and characterizing ASD.

The crucial difficulties researchers are encountering in this research are due to the underlying biological heterogeneity of ASD. The candidate biomarkers generally fail to match the clinical categories as defined in the Diagnostic and Statistical Manual of Mental Disorders, which do not reflect the evidence that heterogeneity in the expression of the ASD condition spans along numerous phenotypic dimensions, overlapping with those of other conditions and of the general population [17].

Brain imaging in ASD, despite its limited diagnostic utility, has already demonstrated a fundamental role in the ASD characterization, through the in vivo observation of the brain involvement in the disorder [18]. Thanks to the non-invasive and non-harmful nature of magnetic resonance imaging (MRI) examinations, many studies have been conducted to reveal volumetric abnormalities of the brain using structural MRI (sMRI), structural connectivity through diffusion-weighted imaging studies, disease-related functional activation patterns and functional connectivity via functional MRI (fMRI) and resting-state fMRI (rs-fMRI),

Given the complexity of this scenario, comprehensive answers to the many open issues can be attempted only by working on larger datasets of better characterized subjects, including detailed description of clinical phenotypes, complete assessment with neuroimaging, and genetic, environmental and immunologic profiles. Collaborative multicenter studies are thus necessary,

since data sharing reduces the difficulties of acquiring neuroimaging data of enough subjects so as to draw meaningful conclusions.

The difficulties of working with large datasets have motivated the need for powerful computational approaches for handling large amounts of data, and for the development of new algorithms to analyze multicenter and multimodal data. Multicenter data require harmonization, whereas new strategies have to be implemented for extracting subtle information and interrelationships that are hidden in multimodal data. Moreover, accounting for the possibility that some information is missing for a subset of subjects within the collected data, robust methods for handling analyses in case of missing data have to be fully developed and optimized with respect to the specific underlying scientific question.

Recent developments in neuroinformatics have provided neuroscientists with suitable tools for an efficient handling of the imaging and clinical data of subjects, especially in large-scale multicenter and multimodal studies, making imaging data available in a variety of formats, and ensuring that all information is stored as anonymous data. Among the successful examples of neuroinformatics tools developed to this purpose are the Biomedical Informatics Research Network (BIRN) [19], XNAT Central [20], the neuroimaging informatics tools and resources clearinghouse (NITRC) [21], and the neuroimaging suite (COINS), that includes web-based tools to manage studies, subjects, imaging, and clinical data [22]. Virtual laboratories for data processing are also becoming available to the neuroimaging community, as provided by the FP7-funded neuGRID2 through a portal that allows users to access to powerful grid/cloud-based computing facilities to process public or proprietary data repositories [23].

To tackle the issues of the complex scenario of ASD, the ARIANNA project aims at developing a novel research environment for studies of this disorder through neuroimaging (https://arianna.pi.infn.it). The innovative ARIANNA approach consists in disjoining the planning of the experiment and the data acquisition from the data analysis. The first phases will be conducted by Psychiatry researchers and MRI scientists, the latter will be performed outside the Psychiatry

research team that designed the study, and shall be transmitted to the ARIANNA Data Analysis Team (DAT). The DAT members have expertise in both standard and innovative data processing techniques and manage adequate computing resources to efficiently carry out demanding analyses of neuroimaging data. Therefore, the ARIANNA project will release to the Scientific Community a web-based service for the analysis of neuroimaging data and related subjects' information. This service will be freely accessible to Psychiatry researchers (i.e. the end users), which will transmit specific data analyses to the DAT. To stimulate the use of the ARIANNA research environment by the International community of researchers on ASD, dedicated "Calls for data" will be launched by the interdisciplinary team of ARIANNA developers, which include Psychiatrists, Physicists, Computer Scientists and experts in Legal Informatics.

This study provides a comprehensive description of the ARIANNA infrastructure, along with the successful functionality tests on external users' access, imaging data uploading and storage, data selection and data analyses on a dedicated computing farm equipped with GPUs. To test the efficiency of the available computing resources, we run test both on a proprietary data sample of 152 subjects and on the publicly available sample of 1112 subjects of the Autism Brain Imaging Data Exchange (ABIDE) project (http://fcon_1000.projects.nitrc.org/indi/abide/).

2-Methods

2.1 ARIANNA platform architecture

A schematic overview of the ARIANNA platform is shown in Figure 1. The entire infrastructure is hosted at the INFN Pisa scientific computing center (arianna.pi.infn.it). A dedicated web interface enables both data providers and data analysts to access the system through a web browser. Data providers can easily upload both clinical data and brain scans (both structural and functional MRI acquisitions). As soon as the data are uploaded, an automated validation procedure is launched in background to: i) verify the validity of the data; ii) assure that no sensitive data are included in the

DICOM header files; iii) convert the DICOM files in a single volume file in NIFTI format; iiii) store the inserted data in an SQL database.

Data analysts can interrogate the database through the web interface according to specific criteria. For this purpose, a graphical interface is available to execute queries on the stored data. A list of selected subjects becomes available along with the corresponding variables so that the data analysts are able to redirect the data to the processing workflows and to the statistical tools available in the ARIANNA computing resources. Finally, data analysts are able to retrieve the generated results and interpret them to answer scientific questions and test working hypotheses.

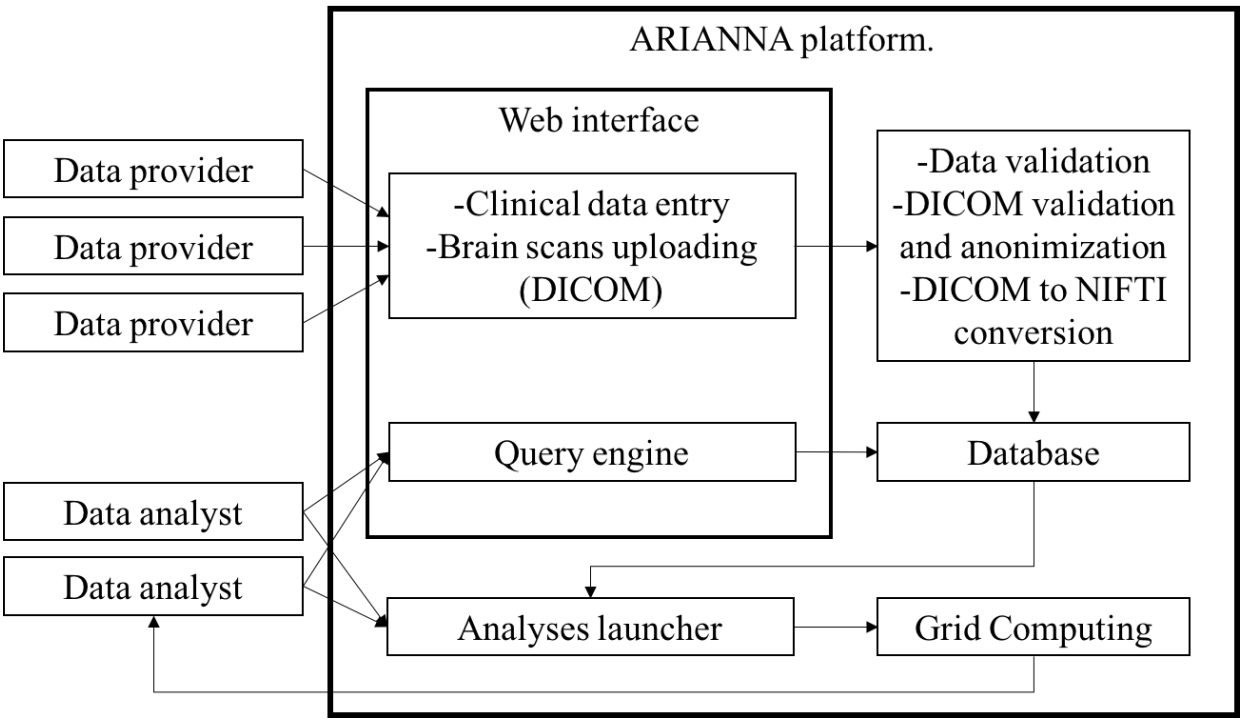


Figure 1: Diagram of the ARIANNA platform

2.2 Web interface

The web interface has been implemented with Drupal (www.drupal.org). Drupal is a content management software used to build websites and applications. It is open source software distributed

under the terms of the GNU General Public License (GPL) that is designed to provide functionalities to collaborative and innovative initiatives. The flexibility and modularity of Drupal make it particularly suited for the development of a collaborative environment such as ARIANNA.

The ARIANNA web interface provides different functionalities and access permissions to various types of users (platform administrators, data providers and DAT members).

2.3 Data entry

The data entry system requires a number of demographic and clinical variables characterizing each patient. In the first place, gender and age at MRI are necessary. Secondly, the mandatory clinical variables for patients are the intelligence quotient (IQ), assessed according to well-standardized tests, and the Autism Diagnostic Observation Schedule (ADOS) total score. The latter consists in an evaluation of autistic symptoms and of emotional/behavioral problems and the data provider can choose to insert ADOS-G [24] or ADOS-2 [25]. The ADOS is a semi-structured, standardized assessment of communication, social interaction, play, and restricted and repetitive behaviors for use with individuals with possible ASD. Moreover, the presence of epilepsy, medical/psychiatric comorbidities and medications has to be specified. The level of expressive language and the fields regarding the Child Behavior Check List (CBCL 1½-5 and CBCL 6-18) [26] can be optionally filled in. The CBCL is a parent-report screening questionnaire designed to assess the child's emotional, behavioral, and social problems. The data provider can specify the diagnosis of ASD according to the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition Text Revision (DSM-IV-TR) [27] criteria or the DSM-5 [1].

Regarding the control subjects, the data entry system requires the variables gender, age at MRI, motivation for the exam and IQ level with the specification of the type of standardized test used.

The ARIANNA platform accepts the upload of the structural MRI (sMRI), the Diffusion Tensor Imaging (DTI) or the resting-state functional MRI (rs-fMRI) scans. One or more of these

acquisitions can be uploaded and they can also be inserted at different time points, in order to allow longitudinal studies.

The Drupal web interface provides a user-friendly environment for data entry. Drop-down and multiple-choice menus simplify the input of categorical variables, while live controls check if the inserted numerical scores fall within the proper ranges. An example of data entry web interface is shown in Figure 2.

2.4 Imaging Data validation and protection

A specific procedure has been implemented to validate neuroimaging data. Data providers specify the type of acquisition and upload a compressed folder containing the 2D DICOM files of the correspondent MR sequence. Once uploaded, a validation process is launched in background to verify the integrity of the data, to retrieve useful information from the DICOM header files (age at scan, pixel size, slice thickness), and to anonymize all the DICOM attributes that may potentially result in leakage of information that could identify individuals.

DICOM files validation, anonymization and information retrieval from header files are performed by taking advantage of the functionalities of the DCMTK (DICOM ToolKit) libraries (http://dicom.offis.de/dcmtk.php.en).

The anonymization task is achieved according to the guideline provided in the DICOM standard description in the Security and System Management Profiles document[1] as detailed in Annex E (Attribute Confidentiality Profiles). In particular, by means of the DICOM modifier command (dcmodify) provided by the DCMTK library, each DICOM header is processed by a bash script which erases the content of all the DICOM attributes that could contain individually identifiable information. The verified and anonymized DICOM files are then converted to a single NIFTI file that contains the entire brain volume which is then saved on the storage elements of the ARIANNA computing center. Indeed, the NIFTI format has become the most widespread brain acquisition

[1] http://dicom.nema.org/medical/dicom/current/output/pdf/part15.pdf

format in neuroimaging research. The validated clinical data and the location on the platform of the correspondent imaging files are then saved in an SQL database, ready to be retrieved and used by the DAT. The DAT performs an initial quality check of the data to assure that the uploaded data are not affected by motion or other artifacts which could affect the quality of the results. Data protection is relevant for the type of data and involved subjects as well as for the nature of the system, the aim of which is providing a data service for sharing the information among research centers. As such, it is fundamental that data are treated in accordance with the Italian Law and EU regulations. Further to these propositions, data will be handled according to rules included in the regulations on protection of personal data (*Codice sulla protezione dei dati personali*, D. Lgs. n. 196/2003), which, in the Article 26, define the rules and methodologies for the treatment of sensitive data, including those for health research purposes. It is worthwhile to highlight that, in addition to individual consent, permission to the Privacy Guarantor (*Garante della Privacy*) is requested.

Such an authorization is however conditioned to the approval of the Regional Ethics Committee in charge of the project or the scientific activity for which the processing of personal data is necessary. Data on the ARIANNA platform will be processed in order to make available only the information strictly required for achieving the scientific objectives of the project, ensuring anonymization/cancellation of all other information. The modification of rules on data privacy, subject to intervention at various levels by the platform developers so as to maintain its compliance with regulations, has to be constantly monitored by the project partners [28]. Additionally, the system provides security measures to prevent unauthorized access to the data repository and ensures an appropriate access to the platform and resulting data handling which are based on the different user profiles (researchers, data providers, DAT members, etc.).

2.5 ARIANNA database query engine

The ARIANNA database, constituted of the data uploaded by the participating data providers, can be interrogated through a dedicated web interface (Figure 3). A selection of the available subjects can be made according to diverse criteria including gender, ranges of age, ranges of symptoms' gravity, IQ levels, and availability of specific brain scan modality.

The outcome of the query engine is a list of the selected subjects along with the available correspondent variables and the location of the neuroimaging files on the storage elements. In this way, the DAT members can easily redirect the desired subset to the grid computing environment. Moreover, all the clinical variables are available for the subsequent statistical analyses, which can be carried out once neuroimaging outcomes have been extracted.

2.6 Analyses launcher and grid computing

ARIANNA computing and storage resources are located at the INFN Pisa scientific computing center. The computing center was established as a Tier2 for the LHC CMS experiment in the early 2000s. Since then, the infrastructure has evolved and currently it also hosts the National center for Theoretical Physics Community and industrial collaboration based on Computing Fluid Dynamics applications (a strong collaboration exists with the Aerospace Engineering Department of the Pisa University).

In 2013, the first farm equipped with GPUs was installed in the data center. Since then, many research groups have used this technology for their own purposes, and today a dedicated farm performs quantum chromodynamics simulations on behalf of an international theoretical physics collaboration.

For neuroimaging purposes, GPUs can reduce the computational times of algorithms with respect to traditional CPUs, thanks to their high-performance, data-parallel architecture.

Therefore, the ARIANNA DAT is able to process neuroimaging datasets – previously selected by a query on the database – on the dedicated farm. The farm currently consists of a single server equipped with:

- CPUs: 2x 10 cores Intel Xeon E5-2640v4 @2.40 GHz;
- RAM: 64 GB;
- GPUs: 4x nVidia Tesla K80, with 2x GPUs Tesla GK210, 24 GB RAM and 2496 CUDA cores each;
- all the neuroimaging software required by the community.

Alongside this computation facility, users have 72 TB of total disk space for the storage of the processed images. The storage has been set up with NFSv4 and it has been designed to guarantee secure data handling.

Access to the computing resources of the center is guaranteed by a user interface. This user interface is intended to be used strictly for the jobs' submission, implemented via the LSF scheduler and based over a batch queue.

2.7 Data preprocessing and analysis pipelines available in ARIANNA

The ARIANNA project provides the Scientific Community with a web-based service for the analysis of neuroimaging data and related subjects' information. Specific "Calls for data" are periodically launched by the ARIANNA developers so as to stimulate the collaboration among researchers of different Institutions in order to address uncovered issues in ASD research. Once the external users that joined the project and the ARIANNA team have discussed and planned a set of analyses on the shared MRI dataset, the DAT conducts the entire analysis, including preprocessing of brain data, quality control of the extracted measurements, statistical tests and data mining procedures. To fulfill these tasks, the ARIANNA platform is equipped with general-purpose methods and tools for neuroimaging data preprocessing and analysis, as detailed below.

2.7.1 Preprocessing pipelines

Three of the most widespread toolboxes for the processing of MRI data are available to the DAT on the ARIANNA computing resources: Statistical Parametric Mapping (SPM) software version

SPM12 (www.fil.ion.ucl.ac.uk/spm/), FSL software version 5.0 (fsl.fmrib.ox.ac.uk/fsl/) and FreeSurfer software version 5.3 and 6.0 (surfer.nmr.mgh.harvard.edu). These toolboxes provide workflow and functions to preprocess and extract significant measurements from structural MRI (both T1-weighted and DTI) and fMRI scans. For example, some of the preprocessing pipelines that the DAT can easily apply are the Voxel-Based Morphometry (VBM) preprocessing in SPM12, the FreeSurfer recon-all tool for cortical parcellation (according to the Desikan-Killiany, the Desikan-Killiany-Tourville and the Destrieux atlases), and the FreeFurfer and FSL-FIRST methods for subcortical structure segmentation. Moreover, the quality control of the extracted features is carried out by means of the scripts provided by the ENIGMA collaboration, that allow a visual inspection of the outcomes by axial, sagittal and coronal views for the parcelled brain cortex (and analogously for the segmentation of subcortical structure) of each subject.

2.7.2 Analysis methods

Subsequent to the preprocessing step, the typical analyses the ARIANNA team can perform include VBM and FreeSurfer-based statistical comparisons of the structural features of both the cortex and inner structures on each data sample, as well as on the entire sample, accounting for the variability across different data acquisition sites. To handle multi-site information, a site-specific ID will be assigned to each Institute in order to address potential effects on the analysis that depend on the site of provenance. The ARIANNA platform enables both seed-based correlation and ICA-based analysis for detection of fMRI resting signal correlations and multiple functional networks identification. Regarding DTI, ARIANNA provides the DAT with voxel-wise statistical analyses of fractional anisotropy maps and tools for automatic reconstruction of major white-matter pathways.

Data mining techniques such as Support Vector Machine (SVM), random forest and logistic regression are available to perform multivariate analyses to identify brain regions specifically involved in ASD. In addition, innovative analysis techniques, including those based on machine-

learning, can be applied to more homogeneous subgroups of subjects, grouped according to narrower age ranges and IQ values, to detect a more robust neuroimaging-based biomarker.

The analysis methods recently implemented by our group on data samples of young children with ASD [29–36] can be fully exploited on the wider samples collected within the ARIANNA project.

Once the DAT team has finalized the post processing steps, the users are provided with statistical summaries, significance maps and main results to be jointly discussed and to be eventually directly included in scientific reports and manuscripts for peer-reviewed journals.

3-Results

3.1-Proof of concept analyses

Functionality tests on the ARIANNA platform have been performed on the data sample (a group of 38 male and 38 female preschoolers with ASD and a group of 76 age- and sex-matched control subjects) recently analyzed by some of the authors with a voxel-based machine-learning technique [30]. Participants were recruited at the ASD Unit of IRCCS Stella Maris Foundation (Pisa), a tertiary hospital and research university in Italy. The clinical information and T1-weighted MRI brain scans of the 152 subjects were uploaded through the web-interface in order to test the entire workflow of data entry, data validation, brain scans anonymization, conversion to the NIFTI format and storage in the SQL database.

Then a query was performed through the query engine to identify and retrieve the subgroups of ASD male subjects, ASD female subjects, control male subjects and control female subjects from the storage elements. The retrieved data were then redirected to the computing resources, and subcortical structures segmentation was performed with FreeSurfer 5.3.0 (recon-all workflow).

The results were then downloaded on local resources to perform statistical analyses of the volumetric measurements extracted for each subcortical region of interest. An example of comparisons of volumes of the subcortical regions' left and right caudate nucleus, already known to

be potentially altered in ASD with respect to controls [37][38][39], are shown in Figure 4. Significantly, caudate nuclei show a statistically significant increase in terms of segmented volume. In particular, the volumes of caudate nuclei measure (left) 3.3 ± 0.5 ml for the control and 3.6 ± 0.5 ml for the ASD population and (right) 3.5 ± 0.5 ml for the control and 3.8 ± 0.5 ml for the ASD population; they differ significantly on the ANOVA test when sex and Total Brain Volume are considered as covariates with p=0.024 and for the left and p=0.025 for the right caudate nucleus, respectively.

In order to demonstrate the potentiality of the ARIANNA platform and its computing resources, we carried out an identical analysis on an ABIDE dataset [40], a large public dataset which consists of 1112 subjects, including 539 individuals with ASD and 573 control subjects aged between 7 and 64 years.

Moreover, we also evaluated the ARIANNA platform in terms of efficiency of the computing resources both on the already described data sample of 152 children and on the ABIDE cohort. The recon-all mean running time was 6.85 hours/case by using the ARIANNA CPUs and was reduced to 5.76 hours/case by exploiting the GPUs parallel computation on the computing node. It has to be considered that the recon-all workflow still lacks optimization for GPUs parallel computing. Other algorithms that ARIANNA planned to implement and use on daily basis, e.g. machine learning algorithms based on Deep Learning, can be boosted by GPUs computing and reduce computing times by factors up to dozens of orders of magnitude [41].

4-Conclusions

The functionality of the ARIANNA platform in terms of external users' access, imaging data uploading and storage has been fully demonstrated. The research environment is now ready to serve the target Scientific Community of researchers working on ASD, enabling collaborative research through neuroimaging data analysis (https://arianna.pi.infn.it). Data processing and analysis is carried out by the team of ARIANNA data analysts (the DAT), which will apply consistent analysis

protocols to multi-site data collected within the project in a collaborative effort. This strategy fosters the concepts of data recycling and data repurposing to respond to different or additional scientific questions with respect to the ones for which the data were originally acquired. Adequate computing resources and analysis protocols are available in the ARIANNA platform to shed new light on ASD through neuroimaging studies.

The working environment proposed in the ARIANNA project can be replicated to meet the requirements of other research communities working on different disease conditions. Moreover, it can be extended to accommodate other data acquisition modalities (e.g. magnetic resonance spectroscopy, EEG, PET) in addition to subjects' clinical and pharmacological history, and patients' genetic profile to enable investigation on other brain diseases and enlarge the community of involved Neuroscientists.

Funding

This work has been supported by the Tuscany Government (PAR-FAS 2007-2013, Bando FAS Salute 2014) through the ARIANNA Project (C52I16000020002).

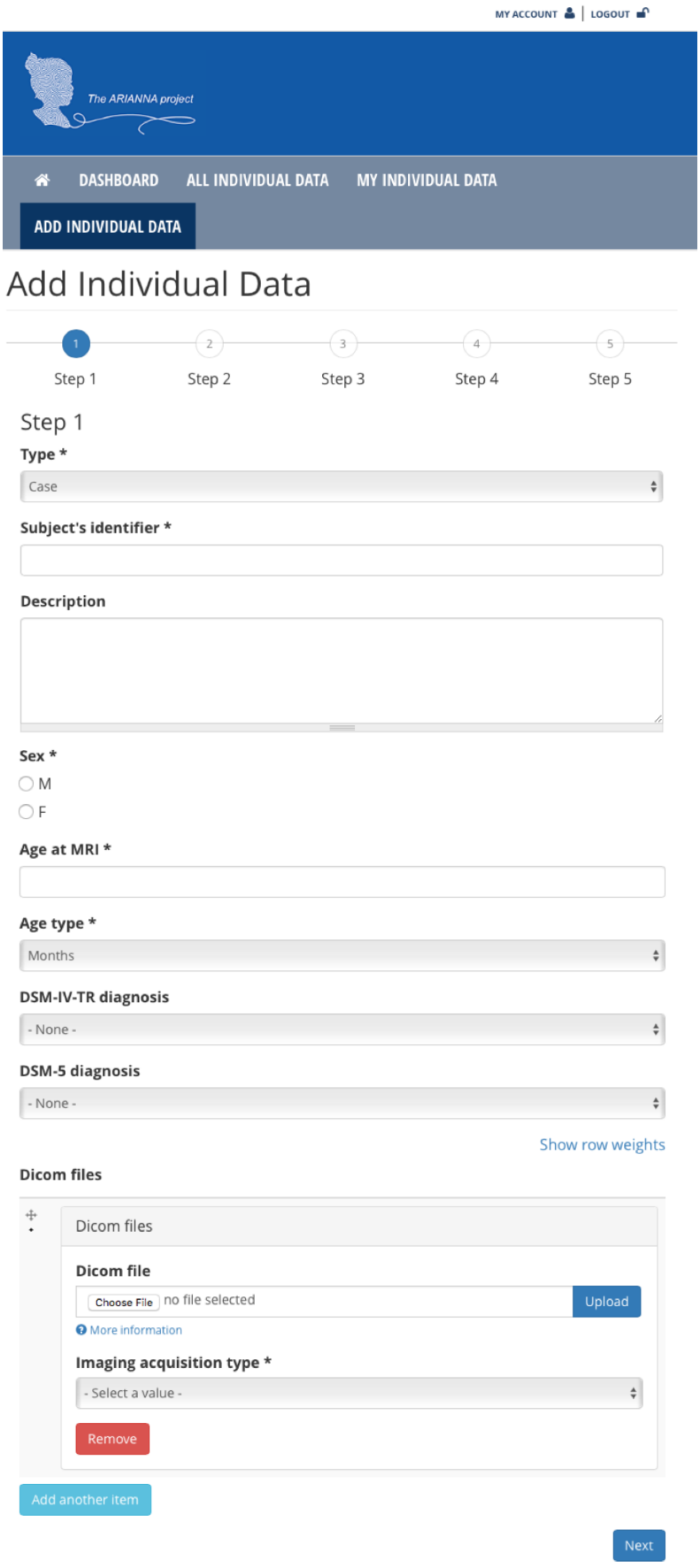


Figure 2: ARIANNA data entry web interface

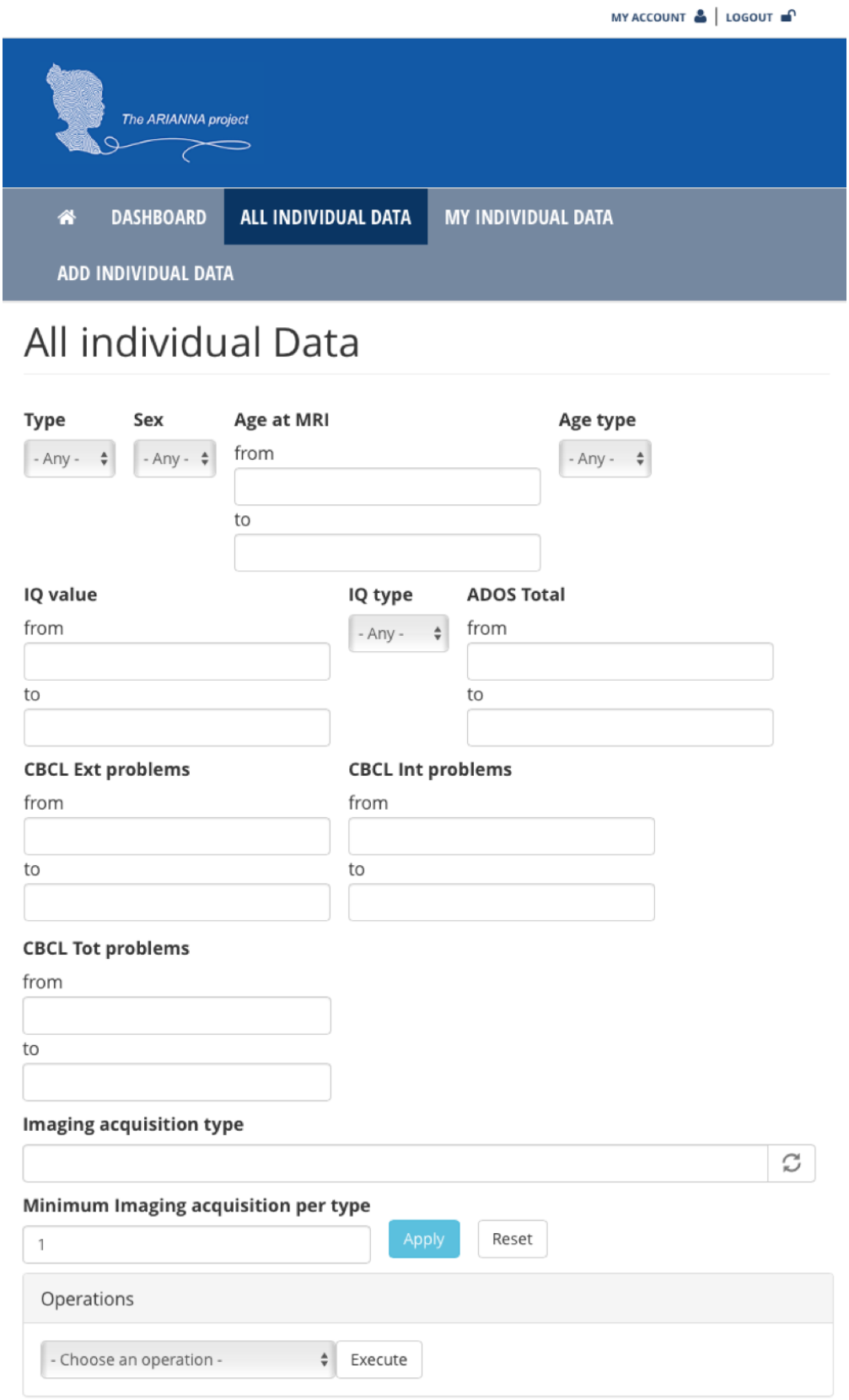


Figure 3: ARIANNA Query interface

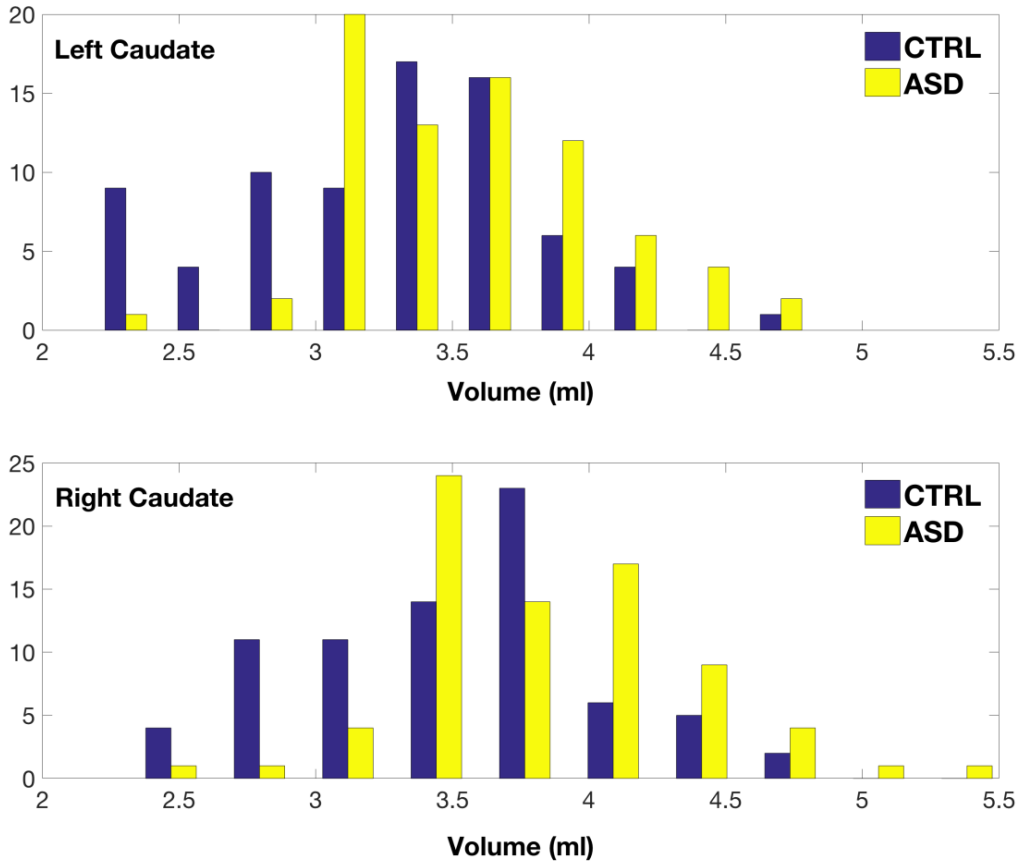


Figure 4: Distributions of the volumes of left and right Caudate Nucleus, segmented by Freesurfer recon-all algorithm for both control (blue) and ASD (cohorts)

References

[1] Association AP, others. Diagnostic and Statistical Manual of Mental Disorders, (DSM-5) 2014.

[2] Christensen DL, Baio J, Braun KVN, Bilder D, Charles J, Constantino JN, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. MMWR Surveill Summ 2016;65:1–23. doi:10.15585/mmwr.ss6503a1.

[3] Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA 2014;311:1770–7. doi:10.1001/jama.2014.4144.

[4] Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. Pediatrics 2010;125:e17-23. doi:10.1542/peds.2009-0958.

[5] Calderoni S, Billeci L, Narzisi A, Brambilla P, Retico A, Muratori F. Rehabilitative interventions and brain plasticity in autism spectrum disorders: Focus on MRI-based studies. Front Neurosci 2016;10:139. doi:10.3389/fnins.2016.00139.

[6] Zwaigenbaum L, Bauman ML, Stone WL, Yirmiya N, Estes A, Hansen RL, et al. Early Identification of Autism Spectrum Disorder: Recommendations for Practice and Research.

Pediatrics 2015;136:S10–40. doi:10.1542/peds.2014-3667C.
[7] Jones RM, Lord C. Diagnosing autism in neurobiological research studies. Behav Brain Res 2013;251:113–24. doi:10.1016/j.bbr.2012.10.037.
[8] Walsh P, Elsabbagh M, Bolton P, Singh I. In search of biomarkers for autism: scientific, social and ethical challenges. Nat Rev Neurosci 2011;12:603–12. doi:10.1038/nrn3113.
[9] Scherer SW, Dawson G. Risk factors for autism: translating genomic discoveries into diagnostics. Hum Genet 2011;130:123–48. doi:10.1007/s00439-011-1037-2.
[10] Schwarz E, Guest PC, Rahmoune H, Wang L, Levin Y, Ingudomnukul E, et al. Sex-specific serum biomarker patterns in adults with Asperger's syndrome. Mol Psychiatry 2011;16:1213–20. doi:10.1038/mp.2010.102.
[11] Yap IKS, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK. Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. J Proteome Res 2010;9:2996–3004. doi:10.1021/pr901188e.
[12] Elder LM, Dawson G, Toth K, Fein D, Munson J. Head circumference as an early predictor of autism symptoms in younger siblings of children with autism spectrum disorder. J Autism Dev Disord 2008;38:1104–11. doi:10.1007/s10803-007-0495-9.
[13] Muratori F, Calderoni S, Apicella F, Filippi T, Santocchi E, Calugi S, et al. Tracing back to the onset of abnormal head circumference growth in Italian children with autism spectrum disorder. Res Autism Spectr Disord 2012;6:442–9. doi:10.1016/j.rasd.2011.07.004.
[14] Ecker C, Murphy D. Neuroimaging in autism--from basic science to translational research. Nat Rev Neurol 2014;10:82–91. doi:10.1038/nrneurol.2013.276.
[15] Pierce K, Conant D, Hazin R, Stoner R, Desmond J. Preference for geometric patterns early in life as a risk factor for autism. Arch Gen Psychiatry 2011;68:101–9. doi:10.1001/archgenpsychiatry.2010.113.
[16] Billeci L, Narzisi A, Campatelli G, Crifaci G, Calderoni S, Gagliano A, et al. Disentangling the initiation from the response in joint attention: an eye-tracking study in toddlers with autism spectrum disorders. Transl Psychiatry 2016;6:e808. doi:10.1038/tp.2016.75.
[17] Volkmar FR, State M, Klin A. Autism and autism spectrum disorders: diagnostic issues for the coming decade. J Child Psychol Psychiatry 2009;50:108–15. doi:10.1111/j.1469-7610.2008.02010.x.
[18] Retico A, Tosetti M, Muratori F, Calderoni S. Neuroimaging-based methods for autism identification: A possible translational application? Funct Neurol 2014;29:231–9.
[19] Keator DB, Grethe JS, Marcus D, Ozyurt B, Gadde S, Murphy S, et al. A National Human Neuroimaging Collaboratory Enabled by the Biomedical Informatics Research Network (BIRN). IEEE Trans Inf Technol Biomed 2008;12:162–72. doi:10.1109/TITB.2008.917893.
[20] Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The extensible neuroimaging archive toolkit. Neuroinformatics 2007;5:11–33. doi:10.1385/NI:5:1:11.
[21] Buccigrossi R, Ellisman M, Grethe J, Haselgrove C, Kennedy DN, Martone M, et al. The Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC). AMIA Annu Symp Proc 2008:1000.
[22] Scott A, Courtney W, Wood D, de la Garza R, Lane S, King M, et al. COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets. Front Neuroinform 2011;5:33. doi:10.3389/fninf.2011.00033.
[23] Redolfi A, McClatchey R, Anjum A, Zijdenbos A, Manset D, Barkhof F, et al. Grid infrastructures for computational neuroscience: the neuGRID example. Future Neurol 2009;4:703–22. doi:10.2217/fnl.09.53.
[24] Lord C, Risi S, Lambrecht L, Cook, Jr. EH, Leventhal BL, DiLavore PC, et al. The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. J Autism Dev Disord 2000;30:205–23. doi:10.1023/A:1005592401947.
[25] McCrimmon A, Rostad K. Test Review: Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module. J Psychoeduc Assess 2014;32:88–92.

doi:10.1177/0734282913490916.
[26] Achenbach TM, McConaughy SH, Ivanovaa MY, Families LARRC for CY and. Manual for the ASEBA Brief Problem Monitor TM (BPM) 2011:13.
[27] Eroglu S., Toprak S., Urgan O, MD, Ozge E. Onur, MD, Arzu Denizbasi, MD, Haldun Akoglu, MD, Cigdem Ozpolat, MD, Ebru Akoglu M. DSM-IV Diagnostic and Statistical Manual of Mental Disorder. vol. 33. 2012. doi:10.1073/pnas.0703993104.
[28] Romano F, Boschi L, Spinicci E, Mercatali P. Automatic Translation from textual representation of laws to formal models through UML. 18th Annu. Conf. Leg. Knowl. Inf. Syst. (JURIX 2005), n.d.
[29] Giuliano A, Saviozzi I, Brambilla P, Muratori F, Retico A, Calderoni S. The effect of age, sex and clinical features on the volume of Corpus Callosum in pre-schoolers with Autism Spectrum Disorder: a case-control study. Eur J Neurosci 2017. doi:10.1111/ejn.13527.
[30] Retico A, Giuliano A, Tancredi R, Cosenza A, Apicella F, Narzisi A, et al. The effect of gender on the neuroanatomy of children with autism spectrum disorders: a support vector machine case-control study. Mol Autism 2016;7:5. doi:10.1186/s13229-015-0067-3.
[31] Valvo G, Baldini S, Retico A, Rossi G, Tancredi R, Ferrari AR, et al. Temporal lobe connects regression and macrocephaly to autism spectrum disorders. Eur Child Adolesc Psychiatry 2016;25:421–9. doi:10.1007/s00787-015-0746-9.
[32] Retico A, Gori I, Giuliano A, Muratori F, Calderoni S. One-class support vector machines identify the language and default mode regions as common patterns of structural alterations in young children with autism spectrum disorders. Front Neurosci 2016;10. doi:10.3389/fnins.2016.00306.
[33] Giuliano A, Gori I, Muratori F, Saviozzi I, Oliva P, Tancredi R, et al. Machine learning techniques implemented on structural MRI features at different spatial scales for preschoolers with autism spectrum disorders. Phys Medica 2016;32:128. doi:10.1016/j.ejmp.2016.01.443.
[34] Gori I, Giuliano A, Muratori F, Saviozzi I, Oliva P, Tancredi R, et al. Gray Matter Alterations in Young Children with Autism Spectrum Disorders: Comparing Morphometry at the Voxel and Regional Level. J Neuroimaging 2015;25:866–74. doi:10.1111/jon.12280.
[35] Retico A, Tosetti M, Muratori F, Calderoni S. Neuroimaging-based methods for autism identification: A possible translational application? Funct Neurol 2014;29:231–9.
[36] Calderoni S, Retico A, Biagi L, Tancredi R, Muratori F, Tosetti M. Female children with autism spectrum disorder: An insight from mass-univariate and pattern classification analyses. Neuroimage 2012;59:1013–22. doi:10.1016/j.neuroimage.2011.08.070.
[37] Rojas DC, Peterson E, Winterrowd E, Reite ML, Rogers SJ, Tregellas JR. Regional gray matter volumetric changes in autism associated with social and repetitive behavior symptoms. BMC Psychiatry 2006;6:56. doi:10.1186/1471-244X-6-56.
[38] Langen M, Durston S, Staal WG, Palmen SJMC, van Engeland H. Caudate Nucleus Is Enlarged in High-Functioning Medication-Naive Subjects with Autism. Biol Psychiatry 2007;62:262–6. doi:10.1016/j.biopsych.2006.09.040.
[39] Langen M, Schnack HG, Nederveen H, Bos D, Lahuis BE, de Jonge M V., et al. Changes in the Developmental Trajectories of Striatum in Autism. Biol Psychiatry 2009;66:327–33. doi:10.1016/j.biopsych.2009.03.017.
[40] Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 2014;19:659–67. doi:10.1038/mp.2013.78.
[41] Raina R, Raina R, Madhavan A, Ng AY. Largescale deep unsupervised learning using graphics processors. Int CONF Mach Learn 2009.