RESEARCH CENTRE

Inria Saclay Centre at Université Paris-Saclay

IN PARTNERSHIP WITH:

Centre CEA-Saclay

2024

ACTIVITY REPORT

Project-Team

MIND

Models and Inference for Neuroimaging Data

IN COLLABORATION WITH: Département NEUROSPIN

DOMAIN

Digital Health, Biology and Earth

THEME

Computational Neuroscience and Medicine

Inria

Contents

Project-Team MIND

Creation of the Project-Team: 2022 April 01

Keywords

Computer sciences and digital sciences

A3.1.1. – Modeling, representation

A3.4.1. – Supervised learning

A3.4.2. – Unsupervised learning

A3.4.4. – Optimization and learning

A3.4.5. – Bayesian methods

A3.4.6. – Neural networks

A3.4.8. – Deep learning

A6.2.4. – Statistical methods

A6.2.6. – Optimization

A9.2. – Machine learning

A9.3. – Signal analysis

A9.7. – AI algorithmics

Other research topics and application domains

B1.2. – Neuroscience and cognitive science

B1.2.1. – Understanding and simulation of the brain and the nervous system

B1.2.2. – Cognitive science

B1.2.3. – Computational neurosciences

B2.6.1. – Brain imaging

1 Team members, visitors, external collaborators

Research Scientists

- Philippe Ciuciu [Team leader, CEA, Researcher, HDR]
- Antoine Collas [INRIA, Starting Research Position, from Nov 2024]
- Chaithya Giliyar Radhkrishna [CEA, Researcher]
- Thomas Moreau [INRIA, Researcher]
- Guillaume Staerman [INRIA, Starting Research Position, from May 2024 until Aug 2024]
- Bertrand Thirion [INRIA, HDR]
- Demian Wassermann [INRIA, Senior Researcher, HDR]

Post-Doctoral Fellows

- Ahmad Chamma [INRIA, Post-Doctoral Fellow, from Jul 2024 until Aug 2024]
- Antoine Collas [INRIA, Post-Doctoral Fellow, until Oct 2024]
- Reuben Dorent [INRIA, Post-Doctoral Fellow, from Aug 2024]
- Fatemeh Ghayyem [INRIA, Post-Doctoral Fellow]
- Baptiste Goujaud [INRIA, Post-Doctoral Fellow, from Sep 2024]
- Ce Ju [INRIA, Post-Doctoral Fellow, from Oct 2024]
- Guillaume Staerman [INRIA, Post-Doctoral Fellow, until Apr 2024]
- Matthieu Terris [INRIA, Post-Doctoral Fellow]
- Sheng Wang [CEA, Post-Doctoral Fellow]
- Houssam Zenati [INRIA, Post-Doctoral Fellow]

PhD Students

- Cedric Allain [INRIA, until Mar 2024]
- Pierre-Louis Barbarant [INRIA, from Sep 2024]
- Alexandre Blain [INRIA, from Nov 2024]
- Alexandre Blain [UNIV PARIS SACLAY, until Nov 2024]
- Ahmad Chamma [INRIA, until Jun 2024]
- Thomas Chapalain [INRIA, from Sep 2024]
- Thomas Chapalain [UNIV PARIS SACLAY, until Aug 2024]
- Pierre-Antoine Comby [UNIV PARIS SACLAY]
- Mathieu Dagreou [INRIA, until Nov 2024]
- Merlin Dumeur [UNIV PARIS SACLAY]
- Celestin Eve [UNIV PARIS SACLAY, from Sep 2024]
- Theo Gnassounou [ENS PARIS-SACLAY]

- Gabriela Pilar Gomez Jimenez [INRIA]
- Ambroise Heurtebise [UNIV PARIS SACLAY]
- Louis Jalouzot [INRIA, from Sep 2024 until Sep 2024]
- Alexandre Le Bris [INRIA]
- Julia Linhart [UNIV PARIS SACLAY, from Nov 2024]
- Julia Linhart [ENS PARIS-SACLAY, until Oct 2024]
- Virginie Loison [INRIA]
- Elsa Manquat [AP/HP, until Nov 2024]
- Sonia Mazelet [CEA, from Oct 2024]
- Apolline Mellot [INRIA, until Oct 2024]
- Raphael Meudec [CEA, from Jul 2024]
- Raphael Meudec [INRIA, until Jun 2024]
- Florent Michel [UNIV PARIS SACLAY]
- Denis Hernando Nunez Fernandez [CEA, from Feb 2024]
- Joseph Paillard [Roche, from Apr 2024]
- Caini Pan [CEA, from Dec 2024]
- Ana Ponce Martinez [INRIA, from Oct 2024]
- Angel Reyero Lobo [UNIV TOULOUSE III, from Oct 2024]
- Joaquin Robador [INRIA, from Apr 2024 until Jul 2024]
- Louis Rouillard-Odera [INRIA, until May 2024]
- Victoria Shevchenko [UNIV PARIS - CITE, from Sep 2024]
- Asma Tanabene [CEA, from Feb 2024]
- Alexis Thual [CEA, until Aug 2024]

Technical Staff

- Himanshu Aggarwal [INRIA, Engineer]
- Liza Al-Shikhley [INRIA, Engineer, until Mar 2024]
- Mansour Benbakoura [INRIA, Engineer, from Oct 2024]
- Remi Gau [INRIA, Engineer, from Sep 2024]
- Lionel Kusch [INRIA, Engineer, from Dec 2024]
- Ziad Ahmed Abdelmotagaly Mohamed [INRIA, Engineer, until Feb 2024]
- Yasmin Mzayek [INRIA, Engineer, until Jan 2024]
- Ana Ponce Martinez [INRIA, Engineer, until Sep 2024]
- Soukaina Tichirra [CEA, Engineer, from Dec 2024]
- Jad Yehya [INRIA, Engineer, from Nov 2024]

Interns and Apprentices

- Pierre-Louis Barbarant [INRIA, Intern, from Apr 2024 until Aug 2024]
- Maxime Bertrait [CEA, from Apr 2024 until Nov 2024]
- Ruben Bueno [ECOLE POLY PALAISEAU, Intern, until Mar 2024]
- Celestin Eve [INRIA, Intern, from Apr 2024 until Aug 2024]
- Denis Fouchard [INRIA, Intern, until Feb 2024]
- Mohammed-Hussem Foufa [INRIA, Intern, from Mar 2024 until Sep 2024]
- Louis Jalouzot [INRIA, Intern, from Apr 2024 until Aug 2024]
- Melvine Nargeot [INRIA, Apprentice]
- Lena Oudjman [CEA, Apprentice, from Mar 2024]
- Caini Pan [CEA, from Apr 2024 until Oct 2024]
- Malo Renaudin [INRIA, Intern, from Apr 2024 until Aug 2024]
- Angel Reyero Lobo [UNIV TOULOUSE II JEAN-JAURES, from Apr 2024 until Sep 2024]
- Jad Yehya [INRIA, Intern, from Apr 2024 until Oct 2024]

Administrative Assistant

- Marie Enee [INRIA]

Visiting Scientist

- David Degras [Univ Massachusetts Boston, until Jun 2024]

External Collaborators

- L'Emir Omar Chehab [ENSAE]
- Samuel Davenport [UNIV CALIFORNIE, until Jun 2024]
- David Degras [Univ Massachusetts Boston, from Jul 2024]
- Denis-Alexander Engemann [Roche]
- Remi Flamary [LIX, HDR]
- Alexandre Gramfort [Ministère Economie, HDR]
- Benjamin Lapostolle [CEA, from Apr 2024 until Aug 2024]
- Tuan Binh Nguyen [IMT, until Jul 2024]
- Jade Perdereau [AP/HP]
- Zaccharie Ramzi [CNRS, until May 2024]
- Tom Willems [UNIV BERNE, from Apr 2024 until Jun 2024]

2 Overall objectives

The MIND team, which finds its origin in the PARIETAL team, is uniquely equipped to impact the fields of statistical machine learning and artificial intelligence (AI) in service to the understanding of brain structure and function, in both healthy and pathological conditions.

AI with recent progress in statistical machine learning (ML) is currently aiming to revolutionize how experimental science is conducted by using data as the driver of new theoretical insights and scientific hypotheses. Supervised learning and predictive models are then used to assess predictability. We thus face challenging questions like *Can cognitive operations be predicted from neural signals?* or *Can the use of anesthesia be a causal predictor of later cognitive decline or impairment?*

To study brain structure and function, cognitive and clinical neuroscientists have access to various neuroimaging techniques. The MIND team specifically relies on non-invasive modalities, notably on one hand, magnetic resonance imaging (MRI) at ultra-high magnetic field to reach high spatial resolution and, on the other hand, electroencephalography (EEG) and magnetoencephalography (MEG), which allow the recording of electric and magnetic activity of neural populations, to follow brain activity in real time. Extracting new neuroscientific knowledge from such neuroimaging data however raises a number of methodological challenges, in particular in inverse problems, statistics and computer science. The MINDproject aims to develop the theory and software technology to study the brain from both cognitive to clinical endpoints using cutting-edge MRI (functional MRI, diffusion weighted MRI) and MEG/EEG data. To uncover the most valuable information from such data, we need to solve a large panoply of inverse problems using a hybrid approach in which machine or deep learning is used in combination with physics-informed constraints.

Once functional imaging data is collected the challenge of statistical analysis becomes apparent. Beyond the standard questions (Where, when and how can statistically significant neural activity be identified?), MIND is particularly interested in addressing driving effect or the *cause* of such activity in a given cortical region. Answering these basic questions with computer programs requires the development of methodologies built on the latest research on causality, knowledge bases and high-dimensional statistics.

The field of neuroscience is now embracing more open science standards and community efforts to address the referenced to as "replication crisis" as well as the growing complexity of the data analysis pipelines in neuroimaging. The MINDteam is ideally positioned to address these issues from both angles by providing reliable statistical inference schemes as well as open source software that are compliant with international standards.

The impact of MINDwill be driven by the data analysis challenges in neuroscience but also by the fundamental discoveries in neuroscience that presently inspire the development of novel AI algorithms. The PARIETAL team has proved in the past that this scientific positioning leads to impactful research. Hence, the newly created MIND team formed by computer scientists and statisticians with a deep understanding of the field of neuroscience, from data acquisition to clinical needs, offers a unique opportunity to expand and explore more fully uncharted territories.

3 Research program

The scientific project of MIND is organized around four core developments (machine learning for inverse problems, heterogeneous data & knowledge bases, statistics and causal inference in high dimension, and machine Learning on spatio-temporal signals).

3.1 Machine learning for inverse problems

Participants: P. Ciuciu, A. Gramfort, T. Moreau, D. Wassermann

Inverse problems are ubiquitous in observational science. This necessitates the reconstruction of a signal/image of interest, or more generally a vector of parameters, from remote observations that are possibly noisy and scarce. The link between the parameters of interest and the observations is physics, and is commonly well understood. Yet, the recovery of parameters is challenging as the

problem is often ill-posed due to the ill-conditioning of the forward model. Machine learning is now more frequently used to address such problems, using likelihood-free inference (LFI) to inverse nonlinear systems, or prior learning using bi-level optimization and reinforcement learning to guide the way to collect observations.

3.1.1 From linear inverse problems to simulation based inference

Expected breakthrough: Boosts in MR image quality and reconstruction speed and in spatio-temporal resolution of M/EEG source imaging
Findings: Development of data-driven regularizing functions for inverse problems, as well as deep invertible and cost-effective network architectures amenable to solve nonlinear inverse problems on neuroscience data.

Solving an inverse problem consists in estimating the unobserved parameters at the origin of some measurements. Typical examples are image denoising or image deconvolution, where, given noisy or low resolution data, the objective is to obtain an underlying high-quality image. Inverse problems are pervasive in experimental sciences such as physics, biology or neuroscience. The common problem across these fields is that the measurements are noisy and generally incomplete.

Mathematically speaking, these inverse problem can be formulated as estimating $\mathbf{x}$ from $\mathbf{y}$ $\Gamma(\mathbf{x})\mathbf{b}$. Here, $\mathbf{b}$ is an additive noise and Γ is a (generally non-injective) mapping to a lower-dimensional space. For example, in magneto- and electroenchephalography (M/EEG), Γ is a real linear mapping $\Gamma_{\text{M/EEG}} : \mathbb{R}^N \mapsto \mathbb{R}^M$ and $\mathbf{b}$ is considered white and Gaussian, while in magnetic resonance imaging (MRI), Γ is a complex linear mapping $\Gamma_{\text{MRI}} : \mathbb{C}^N \mapsto \mathbb{C}^M$ and $\mathbf{b}$ is circular complex white Gaussian. Despite the linearity of $\Gamma_{\text{M/EEG}}$ and Γ_{MRI}, estimating $\mathbf{x}$ is a challenging task when the measurements are incomplete, i.e., $M \ll N$ and the problem is ill-posed. This is often the case due to physical limitations on the measurement device (M/EEG) or the acquisition time (MRI). Moreover, the linear Fourier operator Γ_{MRI} only reflects an ideal acquisition process and part of the acquisition artifacts (e.g. B0 inhomogeneity) can be compensated by considering nonlinear models at the cost of estimating additional parameters along with the MR image.

To tackle these inverse problems, using adequate regularization will promote the right structure for the data to be recovered. Over the last decade the members of Mind have proposed state-of-the-art models and efficient algorithms based on sparsity assumptions [85, 137, 108, 91, 138, 121, 120, 106, 111, 107]. MNE is the reference software developed by the team that implements these methods for MEG/EEG data while pysap-mri proposes solvers for MR image reconstruction.

The field is now progressing with novel approaches based on deep learning by either learning the regularization from data in the context of MRI reconstruction [158, 159], or by considering nonlinear models grounded in the physics underlying the data. The team has started to explore this direction using so-called Likelihood-Free Inference (LFI) techniques built on deep invertible networks [161, 119]. A particular application has been on diffusion MRI (dMRI), where we have linked the dMRI signal with physiological tissue models of grey matter tissue [119]. Still in MRI but in susceptibility weighted imaging, another approach [100] has consisted in directly estimating the B0 field map from non-Cartesian k-space data to correct for off-resonance effects in non-Fourier operators Γ_{MRI}. The Mind project will continue along this direction studying nonlinear simulators of imaging data as building blocks. A key aspect of the work proposed is to exploit knowledge on the physics of the data generation mechanisms.

3.1.2 Bi-level optimization

Expected breakthrough: Efficient algorithms to select hyper-parameters and priors for source localisation in MEG and image reconstruction in MRI/fMRI.
Findings: Bi-level optimization solvers exploiting gradients to scale with the large number of samples and hyper-parameters.

In recent years, bi-level optimization – minimizing over a parameter which is itself the solution of another optimization problem – has raised great interest in the machine learning community. Indeed,

many methods in ML reduce to this bi-level framework, typically the problem of hyper-parameter optimization.

In most practical cases, hyper-parameter selection is done using cross-validation (CV), which basically consists in splitting the whole dataset in training and validation sets. The parameters of the method are computed by minimizing a loss function on the training set, and the hyper-parameters are then set by minimizing the loss function on the validation set. This approach is a bi-level optimization problem.

Other instances of such problems can be found in dictionary learning, robust training of neural networks or the use of implicit layers in deep learning. In all these applications, the model or the latent variables are learned by minimizing some loss while the parameters or the dictionary are updated by minimizing a second optimization problem depending on the outcome of the first problem. While theoretical results were produced in the early 70's [99], there are still many challenges related to bi-level optimization that need to be addressed to produce methods that are both theoretically well grounded and computationally efficient. Recently, the members of MIND have published several works related to the subject [68, 69, 80, 92]. We intend to pursue this effort in the following directions.

Stochastic bi-level solvers. Bi-level solvers require the use of the whole training set before doing an update on an outer-level problem: In this sense, they are *full-batch* methods [80]. We propose to study *stochastic* methods for this task, where some improvement on the optimization can be achieved using only a few samples from the training data. Stochastic algorithms are notoriously faster than full-batch methods for large datasets, but are also generally harder to analyse from a theoretical standpoint. In addition to being fast, the proposed algorithm should come with some statistical guarantees. These solvers can have many applications, from stochastic prior learning for inverse problem to hyper-parameters tuning in general machine learning.

Neural Dictionary Learning. Bi-level optimization framework offers a canvas to advance the state of the art in dictionary and prior learning. Indeed, dictionary learning has long been seen as a bi-level optimization problem [136]. Practical algorithms are mainly based on alternate minimization and rarely account for the sub-optimality of each sub-problem. With advances in bi-level optimization and algorithm unrolling [68], we aim at providing efficient and theoretically justified dictionary learning algorithms, that will be able to leverage the technologies of differentiable programming [66, 154].

Deep Equilibrium Models. The use of Deep learning, and in particular unrolled algorithms [115], has introduced a quantum leap in the resolution of inverse problems compared to variational approaches, specifically in terms of computing efficiency and image/signal recovery performance. However, these networks are very demanding in memory for the training, which currently limits their potential. Different methods exist to alleviate this problem both on the modeling (gradient checkpointing, reversible networks) and the implementation side (model parallelism, mixed precision), but come at the expense of larger computational cost. However, a promising research avenue, illustrated by [112], is the use of Deep Equilibrium Models. These models are defined implicitly and amount to unrolling an infinite number of iterations, thereby using much less memory. These implicit layers constitute another instance of bi-level optimization problem and we plan to work on these directions in the near future as a means to address DL image reconstruction in realistic 3D and 4D multi-coil MRI setting, both for structural and functional imaging.

3.1.3 Reinforcement learning for active k-space sampling

Expected breakthrough: New hardware compliant under-sampling patterns in MRI k-space that accelerate anatomical and functional scans while optimizing MR image quality.
Findings: Develop novel principles of active sampling in the reinforcement learning framework which optimizes a sampling policy tightly linked to the reconstructed image quality.

Current under-sampling schemes in MRI allow for shorter scan acquisition times, however at the cost of artifacts in various regions of the reconstructed MR image. These artifacts arise due to uncertainties in some heavily under-sampled regions of the acquired Fourier space (i.e. also called k-space). Modern reconstruction algorithms, with the use of strong priors, either hand-crafted or learned, tend to reduce these uncertainties and behave as if the acquisition is fixed.

To go beyond the state of the art, we argue that there is a need to jointly learn an algorithm that designs the optimal under-sampling pattern in k-space as well as the reconstruction network.

As it can be summarized to learning a sequential decision algorithm, we will rely on reinforcement learning (RL) to build up optimal k-space sampling patterns while enforcing physical constraints on the MRI sequence, as originally proposed in [88, 83, 129].

The k-space acquisition can be modeled by a sampling policy and the rewards for the joint network are based on reconstructed image quality. Under this paradigm, after every fixed scan time, an instantaneous reconstruction can be obtained and the Fourier space uncertainty maps analysed in depth. Based on this, the scan can continue by actively sampling the k-space and enforcing denser samples in regions where uncertainty is larger. In this way, the learned k-space trajectories may become more patient and organ specific. Further, the trajectory can run and lead to instantaneous best results of reconstruction under a given variable scan time budget. These aspects define one of the core directions we will investigate to produce the next generation of state-of-the-art MR data sampling and image reconstruction algorithms. Recent contributions [175, 155] only approach the problem in the Cartesian framework and hence perform 1D variable density sampling along the phase encoding dimension. Given our expertise on non-Cartesian sampling in developing SPARKLING for both for 2D and 3D MR imaging [129, 130, 86], we plan to extend this framework to non-Cartesian acquisition setups while still remaining compatible with hardware constraints on the gradient system. The access to various MRI scanners at CEA/NeuroSpin is necessary and an added advantage to the success of the Mind team.

3.2 Heterogeneous Data & Knowledge Bases

Participants: B. Thirion, D. Wassermann

Inferring the relationship between the physiological bases of the human brain and its cognitive functions requires articulating different datasets in terms of their semantics and representation. Examples of these are spatio-temporal brain images, tabular datasets, structured knowledge represented as ontologies, and probabilistic datasets. Developing a formalism that can integrate all these modalities requires constructing a framework able to represent and efficiently perform computations on high-dimensional datasets as well as to combine hybrid data representations in deterministic and probabilistic settings. We will take on two main angles to achieve this task: on one hand, the automated inference of cross-dataset features, or coordinated representations and on the other hand, the use of probabilistic logic for knowledge representation and inference. The probabilistic knowledge representation part is now well advanced with the Neurolang project. It is yet a long-term endeavor. The learning of coordinated representations is less advanced.

3.2.1 Learning coordinated representations

Expected breakthrough: Process semantic information together with image data to bridge large-scale resources and knowledge bases
Findings: Set up a learning model that leverages heterogeneous data: Images, annotations, texts, and coordinate tables.

Inference is the pathway that leads from data to knowledge. One crucial aspect is that in the context of neuroscience, data comes in different forms: Full texts, images and tables. Annotations may be full texts or simply tags associated with observed images. One challenge is thus to develop automated techniques that learn *coordinated representations* across such heterogeneous data sources.

This learning endeavor rests on several key machine learning techniques: Compression, embeddings, and multi-layer networks. Compression (sketching) consists in building a reduced

representation of some input that leads from large sparse and complex representation to low-dimension ones, while minimizing some distortion criterion. Embedding techniques also create representations, but possibly bias them to enhance some aspects of the data. It thus incorporates prior information on data distribution or the relevance of features. Finally, multi-layer networks create intermediate representation of data that are suitable to achieve a prediction goal. Such representations are rich enough in particular in multi-task settings, where the outputs of the network are multi-dimensional. Following [74], we call such latent data models *coordinated representations*.

Deep learning is well suited to the goal of learning intermediate representations. As an example, we plan to develop a framework that coalesces in one deep learning formulation, the task of estimating brain structures, cognitive concepts, and their relationships.

Brain structures and cognitive concepts will appear as intermediate representations responsible for linking brain activity to observed behavior. However deep learning cannot be considered as a standard means to understand coordinated representations, due to the limited data available, their poor signal-to-noise ratio (SNR) and their heterogeneity. Deep learning needs instead to be adapted by injecting our expertise on statistical structure of the data (see e.g. [117, 140]). Since the challenge is to train such models on limited and noisy data, we will extend our recent work [73] that has developed regularization schemes for deep-learning models: it relies on structured stochastic regularizations (a.k.a. structured dropout). Such approaches are efficient, powerful and can be used in wide settings. We will enhance them with more generic, cross-layer, grouping schemes. Additionally, we will develop two strategies: *i)* aggregation of predictors for variance reduction and stability of the model [117] and *ii)* data augmentation – i.e. learning to augment, based on unlabeled data – to improve the fit with limited data. For this we will consider plausible generative mechanisms.

3.2.2 Probabilistic Knowledge Representation

Expected breakthrough: A domain-specific language (DSL) capable of articulating heterogeneous probabilistic data sources in neuroimaging is a way to relate physiology to cognition.
Findings: Self-optimizing probabilistic solvers for discrete and continuous hierarchical models able to scale for neuroimaging problems.

Neuroscientific data used to infer the relationships between physiology of the human brain and its cognitive function goes well beyond text, image, and tables. Knowledge graphs representing human knowledge, and the ability to encode reasoning strategies in neuroscience are also key to effectively bridge current data-centric approaches and decades-old domain knowledge. A main challenge in performing inferences combining demographic data-centric approaches, imaging measurements, and domain knowledge, is to be able to infer new knowledge soundly and efficiently taking into account the noisy nature of demographic and imaging measurements, and the common open-world assumption of ontologies and knowledge graphs. Such probabilistic hybrid logic approaches are known to be, in general, intractable in the deterministic [71] as well as in the probabilistic case [172]. Nonetheless, there is an opportunity to be seized in identifying tractable segments of probabilistic hybrid logic representations able to solve open neuroscientific questions.

A noticeable opportunity to incorporate all statistical evidence gathered from noisy data into a usable knowledge base is to formalize the inferred relationships into probabilistic symbolic representations [118]. These representations are much better suited to simultaneously handle data across topologies and logic systems, implementing inferential algorithms avoiding the brittleness of deterministic logic as well as causal probabilistic reasoning.

A typical application of such heterogeneous data processing is meta-analytic applications which combine neuroimaging data with results found in the scientific literature. Current tools to perform this task are NeuroSynth or Neuroquery (developed by the team). However, knowledge inferred by such tools is tremendously limited by the expressive power of the language used to query the data. Current meta-analytic tools are able to express queries relating test makers, article annotations, and their relationship with reported brain activations, support propositional logic only. Propositional logic requires the user to explicitly express every desired term with their characteristics and their relationships. Our goal is to extend the inference capabilities of such

applications by leveraging current advances in probabilistic logic languages and embedding them in the Neurolang language. Neurolang enables the encoding of complex knowledge in terms of more expressive queries. Neurolang queries first-order logic segment, $FO^{\neg\exists}$, with a tractable probabilistic extension allowing for high-dimensional and large dataset computations. Such segment of first order logic enables formalising questions such as "what brain areas are most likely reported active in a study specifically when terms related to consciousness are mentioned in such study", hence being able to infer, amongst other tasks, specificity and causality [173] of diverse neuroscience phenomena. To disseminate our results allowing complex expressive searches of massively aggregated diverse data, we will leverage Neurolang. The latter produces a domain-specific language (DSL) for human neuroscience research, while being able to combine imaging data, anatomical descriptions and ontologies. Three main characteristics of the DSL are key to fulfilling this goal: First, it represents neuroimaging-derived information and spatial relationships in a *syntax close to natural language used by neuroscientists* [174]. Second, through a back-end belonging to the Datalog$^{\pm}$ family, it allows *querying ontologies* with the same expressive power as current standards SPARQL and OWL [77]. Finally, we will extend Neurolang to a *probabilistic language able to express graphical models* allowing the implementation of a wide variety of causal inference and machine learning algorithms [76] in high-dimensional settings which are specific to neuroimaging research. In sum, by leveraging *recent advances in deductive database systems* [77] and this novel DSL [174] we will provide a more flexible tool to express and infer knowledge on brain structure-function relationships.

3.3 Statistics and causal inference in high dimension

Participants: A. Gramfort, T. Moreau, B. Thirion, D. Wassermann

Statistics is the natural pathway from data to knowledge. Using statistics on brain imaging data involves dealing with high-dimensional data that can induce intensive computation and low statistical power. Besides, statistical models on large-scale data also need to take potential confounding effects and heterogeneity into account. To address these questions the Mind team will employ causal modeling and post-selection inference. Conditional and post-hoc inference are rather short-term perspectives, while the potential of causal inference stands as a longer-term endeavor.

3.3.1 Conditional inference in high dimension

Expected breakthrough: Obtain statistical guarantees on the parameters of very-high dimensional generalized linear or non-parametric models.
Findings: Develop computationally efficient procedures that allow inference for such models, by leveraging structural priors on the solutions.

Conditional inference consists of assessing the importance of a certain feature in a predictive model, while taking into account the information carried by alternative features. One motivation for using this inference scheme is that brain regions that sustain behavior and cognition are strongly interacting. Taking these interactions into account is critical to avoid confusing correlation with causation in brain/behavior analysis.

Technical difficulties come when the set of explanatory features $\mathbf{X}$ becomes extremely large as frequently met in neuroimaging: Conditioning on many variables (or equivalently, high dimensional variables) is computationally costly and statistically inefficient. The main solutions to date are based either on linear model debiasing [123], as well as simulation-based approaches (knockoff inference [84] or conditional randomization tests [133]). Importantly the latter involves simulating data with statistical characteristics described explicitly (in a parametric family) or implicitly (by samples). There remain two gaps to bridge for these methods: *i)* The computational gap, as the algorithmic complexity of these approaches is typically cubic in the number of samples, unless more efficient generative mechanisms are available; *ii)* the power gap, related to the limited number of available samples. The best solution thus far consists of associating these inference procedures with dimension reduction procedures [147]. The next step is adaptation to more general settings: Conditional inference has been formulated in the linear framework, where it boils down to controlling

that the corresponding coefficient is non-zero, hence it has to be generalized to nonlinear models: Non-parametric models like random forests, then possibly deep networks.

3.3.2 Post-selection inference on image data

Expected breakthrough: Statistical control of false discovery proportion (FDP) for data under arbitrary correlation structure.
Findings: A computationally efficient non-parametric statistical test procedure, and a benchmark against alternative techniques.

Large-scale statistical testing is pervasive in many scientific fields, where high-dimensional datasets are collected and compared with an outcome of interest. In such high-dimensional contexts, false discovery rate (FDR) control [79] is attractive because it yields reasonable power, while providing an explicit and interpretable control on false positives. Yet the FDR rate is the expectation of the FDP. Controlling the FDR does not mean that the FDP is controlled, a distinction that is most often ignored by practitioners. For the sake of scientific reproducibility, there is a need for methods controlling the FDP.

Such an approach has been developed in the context of neuroimaging, namely the *all-resolution inference* framework [162] based on classical multiple correction error control bounds. Yet, the empirical behavior of this method remains to be assessed. Moreover, it has been clearly established that the procedure is over-conservative in some settings [81]. Indeed, it relies on the Simes statistical bound, that is not adaptive to the specific type of dependence for a particular data set. To bypass these limitations, [81] have proposed a randomization-based procedure known as λ-calibration, which yields tighter mathematical bounds that are adapted to the dependency observed in the dataset at hand. It rests on a non-parametric (permutation-based) estimation of the null distribution, leading to tight and valid inference under general assumptions.

In this research axis, we propose to fix some of the open issues with the approach described in [81], namely the choice of a template family to calibrate the error distribution in the permutation procedure. We hope to propose a practical choice for this family to avoid putting the burden of choice on practitioners.

We will characterize by simulations and theoretical arguments the behavior of these error control procedures and develop efficient computational methods for the use of these tools in brain imaging analysis.

3.3.3 Causal inference for population analysis

Expected breakthrough: Provide a reference methodology for causal and mediation analysis in high-dimensional settings.
Findings: Benchmark state-of-the-art techniques and further adapt them to the high-dimensional setting.

Modern health datasets present population characteristics with many variables and in multiple modalities. They can ground prediction and understanding of individual outcomes, using machine learning techniques. Still, heterogeneous variables have complex relationships, making it hard to tease apart each factor in an outcome of interest. Potential outcome theory [164] provides a valuable framework to evaluate the impact of treatment (interventions). Treatment effects can be heterogeneous. In particular, interactions between background and treatment variables have to be considered.

The statistical behavior (consistency and efficiency) under non-parametric models is actively investigated [72, 148]. However, their behavior in high-dimensional settings, when both the number of features and the number of samples are large, is still poorly understood. Our objective is thus to extend the theory and algorithms of causal inference to noisy high-dimensional settings, where the noise level implies that effects sizes are proportionally small, and classic methods often become inefficient and potentially inaccurate due to overfitting. More specifically, we plan to explore the following directions.

Mediation analysis and conditional independence Mediation analysis considers the question of whether a variable z mediates all the effect of another variable x onto a target variable y, *a.k.a.* outcome. It turns out that full-mediation analysis amounts to testing whether $x \perp y|z$ (x is independent from y given z), which is handled by a conditional independence test. When the dimensions of these variables (z in particular, but also x and to some extent y) grow, the underlying statistical inference procedures typically lose power, or even possibly error control. We propose to leverage our experience on such high-dimensional inference problems [93, 146] to set up computationally efficient and accurate solutions to this problem.

Latent variable models and confounders The most important aspect of inferring causal effects from observational data is the handling of confounders, *i.e.,* factors that affect both an intervention and its outcome. For instance, age has a clear impact on brain characteristics as well as on behavior, potentially biasing brain/behavior statistical associations. A carefully designed observational study attempts to measure all important confounders. When one does not have direct access to all confounders, there may exist noisy and uncertain measurements of proxies for confounders. A possible solution to this problem relies on generative modeling, such as Variational Autencoders (VAE) and Generative Adversarial Networks (GANs), to sample the unknown latent space summarizing the confounders on datasets with incomplete information; the seminal work of [135] is promising, but still requires improvements to become usable in realistic settings.

The quest of model selection and validation In the classical potential outcome theory [164], causal effects are determined by both factual and counterfactual outcomes, ground-truth effects can never be measured in an observational study. In the absence of such measures, how can we evaluate the performance of causal inference methods? Addressing this question is an important step for practical problems, in which one has to determine if an effect can safely be considered non-zero, or heterogeneous through a population. We propose to revisit the promising work of [70] analysing in detail the shortcomings of the procedure (regarding both bias and variance), especially when the model becomes high-dimensional.

3.4 Machine Learning on spatio-temporal signals

Participants: P. Ciuciu, A. Gramfort, T. Moreau, D. Wassermann, B.Thirion

The brain is a dynamic system. A core task in neuroscience is to extract the temporal structures in the recorded signals as a means to linking them to cognitive processes or to specific neurological conditions. This calls for machine learning methods that are designed to handle multivariate signals, possibly mapped to some spatial coordinate system (e.g. like in fMRI).

3.4.1 Injecting structural priors with Physics-informed data augmentation

Expected breakthrough: Obtain models with more predictive power when trained on small datasets.
Findings: Efficient data-augmentation strategy tailored to brain signals.

Data augmentation consists of virtually increasing dataset size during learning by applying random, yet plausible, transformations to the input data. In computer vision, this means altering data by applying symmetries, rotations, geometric deformations etc. While such strategies are reasonable for natural or medical images [153], it is still unclear how neural or BOLD signals can be augmented in order to improve prediction performance and robustness.

Some purely data driven strategies have been proposed to augment EEG data using spectral transforms [134] or advanced strategies such as channel, time or frequency masking or phase randomizations [128, 131]. Although dozens of transformations have been considered in the literature to augment EEG signals, it is now apparent that different augmentation strategies should be applied to the data as a function of the prediction task to be handled. For example when considering sleep stage classification or BCI applications, the spatial sampling of electrodes and the

duration of signals varies considerably, with the consequence being that different augmentation parameters and even transformations need to be employed.

In this line of work we will develop algorithms that can quickly identify the relevant augmentation techniques, building for example on [96, 133]. The aim is to provide a system that can automatically learn invariance within a class and across subjects in order to maximize the prediction performance on unseen data. The methodology developed will be relevant beyond neuroscience as long as a family of physics-informed transformations is available for prediction tasks at hand.

3.4.2 Learning structural priors with self-supervised learning

Expected breakthrough: Unveiling the latent structure of brain signals from large datasets without human supervision as well as improving the prediction performance when learning from limited data.
Findings: Self-supervised algorithms for multivariate brain signals.

Self-supervised learning (SSL) is a recently developed area of research that provides a compelling approach for exploiting large unlabeled datasets. With SSL, the structure of the data is used to turn an unsupervised learning problem into a supervised one, called a "pretext task", such as solving Jigsaw puzzles from images [150] or learning how to color gray-scaled images. The representation learned on the pretext task can then be reused for unsupervised data exploration or on a supervised downstream task, with the potential to greatly reduce the number of labeled examples required to train a good predictive model.

In fields like computer vision [150, 142] and time series processing [151], SSL has shown great promise in terms of prediction performance but also in ease of use. Indeed, SSL simplifies model selection and evaluation as it relies on prediction scores and cross-validation, contrarily to unsupervised learning methods like ICA [67].

Recently the team has applied SSL to two large cohorts of clinical EEG data [75] revealing insights on the data without any human supervision. However many challenges remain. For example in MIND, we aim to explore novel SSL strategies applicable to electrophysiology as well as to haemodynamic signals measured with fMRI. As such, our goal is to expand the recent multivariate method we have introduced in the field for the blind deconvolution of BOLD signals in both task-related and resting-state experiments [90].

While rather small networks have been employed so far on EEG data [87, 163] due to limited sets of annotations, the use of SSL tasks opens the possibility to work with much larger labeled datasets, and therefore many more overparametrized models. We aim to explore these directions, hoping to reach a state where pre-trained models could be available for EEG or MEG signals as is presently the case for images or for natural language processing (NLP) tasks.

3.4.3 Revealing spatio-temporal structures with convolutional sparse coding and driven point processes

Expected breakthrough: A novel way to study and quantify temporal dependencies between neural processes, going beyond connectomes based on spectral analysis.
Findings: Temporal pattern finding algorithms that scale to massive MEG/EEG datasets with parallel processing and point-process inference algorithms.

The convolutional sparse linear model is one established unsupervised learning framework designed for signals. Using algorithms known as convolutional sparse coding (CSC), this framework allows for the learning of shift-invariant patterns to sparsely reconstruct a time series. These patterns, also called atoms, correspond to recurrent structures present in the data. While some of our recent advances have improved the computational tractability of these methods [144, 143] and adapted them to neurophysiological data [122, 105, 90], there are still many shortcomings that make them unpractical for applications beyond denoising.

Model validation The main challenge for the evaluation of unsupervised convolutional models comes from current theoretical limitations: What can we guarantee statistically concerning the recovered atoms? Due to their non-convexity, existing algorithms can only guarantee convergence to local minima, which might be sub-optimal. In this setting, it is challenging to quantify if the model parameters are well estimated and if they are actually representative of the signals. In MIND, we aim to develop statistical quantification of the uncertainty associated with such models and in this regard, provide objective selection criteria for the model and its parameters. This topic of research will benefit from our other developments on bi-level optimization (cf. subsubsection 3.1.2) and on FDR control (cf. subsubsection 3.3.2) as well as the expertise of the team members on dictionary learning [144, 141, 143].

Capturing temporal dependencies with point processes Another shortcoming of these models is that they do not capture temporal dependencies between the occurrences of the different atoms. However, neural activity at level of the whole brain is highly distributed. Different brain regions form networks that are characterized by the presence of statistical dependencies in their activity [152]. An interesting question to formulate is how one can model and learn these time dependencies between brain areas from the MEG or EEG recordings using an unsupervised event-based approach such as CSC. One of the approaches considered is based on point processes (PP; [82, 116]). PP are classical tools to study event trains (e.g. sequence of spikes) and to model their dependency structure. We aim here to develop PP-based inference algorithms as a means to capture network effects in different brain areas, but also to quantify how experimental stimuli are affecting the temporal statistics of temporal patterns [152]. To model this latter scenario, we will develop the so-called driven PP. In a second stage, we aim to design fully unsupervised methods to capture the connections between different brain areas leveraging the full temporal resolution of non-invasive electrophysiological signals.

4 Application domains

The four research axes we presented earlier have been thought of in tight interaction with four main applications (large-scale predictive modeling, mapping cognition & brain networks, modeling clinical endpoints, from brain images and bio-signals to quantitative biology and physics).

4.1 Population modeling, large-scale predictive modeling

4.1.1 Unveiling Cognition Through Population Modeling

Linking the human brain's structure and function with cognitive abilities has been a research epicenter for the past 40 years. The sophistication of brain mapping machinery such as MRI, EEG and MEG, has produced a treasure trove of data. Nonetheless, the effect size of the phenomena leading to understanding cognition is often drowned out by noise and inter-individual variability. A main goal of MIND is to simultaneously harness the power of large-scale general purpose datasets, such as the Human Connectome Project (HCP) and the Adolescent Brain Cognitive Development Study (ABCD), as well as small scale high precision ones, such as the Individual Brain Charting (IBC) dataset [157], to understand the link between the human brain's architecture and function, and cognition. PARIETAL's expertise has already been demonstrated in this field. Examples of this include using diffusion MRI (dMRI) to link the brain's macrostructure with language comprehension [89], tissue microstructure with cognitive control [139], functional gradients on the cortical surface [103] to functional territory segregation [156].

MIND project will continue this task by seizing our core methodological developments, described in the previous section, and our global collaborative network of cognitive scientists.

4.1.2 Imaging for health in the general population

Individual differences in brain function and cognition have historically been investigated by studies carried out by individual laboratories having access mainly to small sample sizes. The growing

availability of public large-scale data of epidemiological dimensions curated by dedicated consortia (e.g. UK Biobank) has enabled studying the relationship between cognition and the brain with unparalleled granularity and statistical power. These resources now allow researchers to relate brain signals/images to rich descriptions of the participants including behavioral and clinical assessments in addition to social and lifestyle factors. Machine learning has proven essential when modeling biomedical outcomes from the large-scale and high-dimensional data brought by consortia and biobanks. It is used to to build predictive models of heterogenous biomedial outcomes (cognitive, social, clinical) based on different neuroscientific modalities. Taken together, this facilitates the study of lifestyle and health-related behavior in the general population, potentially revealing risk factors leading to biomarker discovery.

MIND will greatly contribute to this effort by focusing on population modeling as a tool for enhancing the analysis of clinical data and mental health.

4.1.3 Proxy measures of brain health

Clinical datasets tend to be small as sharing of data is not incentivized or institutional and economic resources are missing. As a consequence, the capacity of machine learning to learn functions that relate complex-to-grasp biomedical outcomes to heterogeneous data cannot be fully exploited. This has stimulated growing interest in proxy measures of neurological conditions derived from the general population, such as individual biological aging. One counter-intuitive aspect of the methodology is that measures of biological aging (e.g. via brain imaging) can be obtained by focusing on the age of a person, which is known in advance and is, in itself not interesting as a target. However, by predicting the age, machine-learning can capture the relevant information about aging. Based on a population of brain images, it extracts the best guess for the age of a person, indirectly positioning that person within the population. Individual-specific prediction errors therefore reflect deviations from what is statistically expected [171]. The brain of a person can look similar to brains commonly seen in older (or younger) people. The resulting brain-predicted age reflects physical and cognitive impairment in adults [170, 94, 104] and reveals neurodegenerative processes [132, 113], which could be overlooked without using machine learning.

MIND will extend this line of research in two directions: 1) Assessment of brain age using EEG and non-brain data such as health-records and 2) proxy measures of mental health beyond aging.

4.1.4 Studying brain age using electrophysiology

MRI is not yet available in all clinical situations and certain aspects of brain function are better understood using electrophysiological modalities (M/EEG). Until recently, it was unclear if brain age can be meaningfully estimated from M/EEG. In a recent study [109], we demonstrated, using the Cam-CAN cohort (n 650), that combining MRI and MEG enhanced detection of cognitive dysfunction. The proposed approach not only achieved integration of brain signals from distinct modalities but explicitly handled the absence of MEG or MRI recordings, adapting ideas from [125]. This is key for clinical translation where one cannot afford excluding cases because one modality is missing. In the clinical setting, EEG is predominantly used (and not MEG). Clinical recordings are far noisier than lab EEG and gold-standard source modeling with MRI is rarely done outside the lab. Supported by theoretical analysis and simulations, we found through empirical benchmarks [166] that Riemannian embeddings 1) capture individual head geometry 2) bring robustness to extreme noise and, 3) enable good age prediction from clinical 20-channel EEG (n=1300) with performance close to 306-channel lab MEG.

MIND will extend this line of research by translating EEG-based brain age measures into the hospital setting and probe these in different patient populations in which ageing-related differences in brain structure and function are part of the clinical picture, e.g., neurodevelopmental disorders, postoperative cognitive decline and dementia (cf. subsection 4.3).

4.1.5 Proxy measures of mental health beyond brain aging

Quantitative measures of mental health remain challenging despite substantial research efforts [126]. Mental health, can only be probed indirectly through psychological constructs, e.g. intelligence or

anxiety gauged by valid and statistically relevant questionnaires or structured examinations by a specialist. In practice, full neuropsychological evaluation is not an automated process but relies on expert judgment to confront multiple responses and interpret them in the context of a larger environmental context including the cultural background of the participant. Inspired by brain age, we set out to build empirical measures of mental health [97] by predicting traditional and broadly used psychological constructs such as fluid intelligence or neuroticism in the UK Biobank. Our results have shown that all proxies captured the target constructs and were more useful than the original measures for characterizing real-world health behavior (sleep, exercise, tobacco, alcohol consumption). In the long run, we anticipate that using proxies could complement psychometric assessments by corroborating data and potentially providing more accurate data faster and more efficiently for clinical populations.

MIND will expand this line of research by systematically searching for proxy measures of physical and mental health derived from large clinical population using electronic health records or transcripts from clinical interviews. We will propose a systematic causal analysis (treatment effect size and mediation) to provide a clearer understanding of the relationships between the many variables that characterize mental health. We will study more in detail the impact of general health markers on brain status, as this may well fit much of the *unexplained variance* on brain health.

4.2 Mapping cognition & brain networks

4.2.1 Problem statement

Cognitive science and psychiatry aim at describing mental operations: cognition, emotion, perception and their dysfunction. As an investigation device, they use functional brain imaging, that provides a unique window to bridge these mental concepts to the brain, neural firing and wiring. Yet aggregating results from experiments probing brain activity into a consistent description faces the roadblock that *cognitive concepts and brain pathologies are ill-defined*. Separation between them is often blurry. In addition, these concepts (a.k.a. *psychological constructs*) may not correspond to actual brain structures or systems. To tackle this challenge, we propose to leverage rapidly increasing data sources: text and brain locations described in neuroscientific publications, brain images and their annotations taken from public data repositories, and several reference datasets.

4.2.2 What machine learning can do for neuroscience

Recent works in computer vision [101] or natural language processing [95, 102] have tackled predictions on a large number of classes, getting closer to open-ended knowledge. These approaches, that rely on uncovering some form of relational structure across these classes, in effect capture the semantics of the domain [95], including the similarity structure of the relevant classes and the ambiguities across classes or the multiple aspects of a class. Broadly speaking, these contributions converge to the concept of representation learning [78], i.e. estimating latent factors that *reformulate a learning problem into a new set of input features or output classes* that are more natural for the data and help further analysis. These new tools enable extraction of knowledge, for instance ontology induction, with statistical learning [149]. They are at the root of heterogeneous data integration, such as multi-modal machine learning [74]. The machine learning challenges that we aim to tackle are three-fold:

- Existing multi-modal machine learning techniques have been developed for relatively abundant data, with overall high SNR: text, natural images, videos, sound. These data are most often non-ambiguous, while brain data typically are, due to the low SNR per image and, more crucially, *poor annotation quality*. We propose to tackle this by adapting machine learning solutions to this low-SNR regime: introduction of priors, aggressive dimension reduction, aggregation approaches and data augmentation to reduce overfitting.
- *Leveraging implicit supervisory signals*: While data sources contain lots of implicit information that could be used as targets in supervised learning, there is most often no obvious way to extract it. We propose to tackle this by using additional, ill- or not-annotated data, relying on *self-supervision methods*.

- *Model interpretability*: Our goal is to provide clear assertions on the relationships between brain structures and cognition: the inference should always lead to an updated knowledge base, i.e. updated relationships between concepts pertaining to neuroscience on one hand, psychology on the other hand. Specifically, one should be able to reason about the information extracted within MIND. For this, we will develop dedicated statistical, causal and formal (ontology-based) data analysis schemes.

Associating knowledge engineering with statistical learning to boost cognitive neuroimaging, requires tackling the challenge of multimodal machine learning under noisy conditions with limited data. Doing so, it will capture links between behavior and brain activity, and enable aggregating the information carried by neuroimaging data to redefine and link concepts in psychology and psychiatry.

4.2.3 Perspective taken: combine distributional semantics with brain images

In natural language processing (NLP), *distributional semantics* capture meanings of words using similarities in the way they appear in their environment. We want to adapt these ideas to learn data-driven organizations of psychological concepts. Importantly, applying these techniques solely to the psychology literature merely captures the current status quo of the field. Including brain images is necessary to bring new information.

To link observed cognition to brain activity, two typical statistical learning problems arise: *encoding*, that seeks to describe brain activity from behavior; and *decoding*, that seeks the converse, predicting behavior from brain activity [127]. In addition, statistical modeling of each aspect of the data on its own generates knowledge, typically *spatial decompositions* from resting-state data, and *topic modeling* on descriptions of behavior. The research strategy followed in this proposal is to *combine the different statistical learning problems in a unified framework* to extract core structures from the aggregation of neuroimaging data: on one side brain structures, and on the other side semantic relationships and concepts in psychological sciences.

MIND will in particular publish automated functional meta-analyses to give a systematic assessment of the publicly available data and question the limitations of the current conceptual framework of systems neuroscience as well as of these resources.

4.3 Modeling clinical endpoints

When sufficient data is available, machine learning can be employed to directly model various clinical endpoints (such as diagnosis, drug response, and neuropsychological scores) from brain signals without the need for proxy measures. This approach has the potential to significantly and meaningfully simplify statistical modeling in clinical research. Machine learning facilitates combining heterogeneous input data (different modalities) and does not need high confidence in underlying generative models linking the data to the clinical endpoint. As a consequence, the same class of models can be applied regardless of the endpoint. Its focus is on bounding the approximation error of the endpoint instead of correct parameter estimates. As such, it provides generalizing models that are more robust. Our team has pushed this type of research program through several important collaborations with our European clinical partners using EEG and MRI.

4.3.1 EEG-based modeling of clinical endpoints

Neurological and psychiatric disorders can show complex neurological patterns. Diagnosis is often performed clinically (based on cerebral signs and behavioral symptoms), leading to important variability across doctors. In clinical neuroscience, predicting diagnosis from *brain signals* is therefore a common application. In the clinical context, EEG is an economically viable option that can be applied in a wide array of circumstances. In collaboration with the Salpêtrière Hospital and the Paris Brain Institute (ICM) we have developed and validated an approach for an EEG-based modeling of diagnosis for severely brain injured patients suffering from consciousness disorders (DoC) [110]. Expert-defined features from consciousness studies were rigorously combined using random forest classification. Sensitivity analysis and benchmarks showed robustness across EEG-configurations

(channels, time points), protocols (resting state vs evoked responses), label noise and differences between recording sites. When changes in the signal are more subtle than they are in DoC patients (average power turned out to be one of the strongest stand-alone features) more general approaches are needed.

Our future activities will focus on extending this line of research to other clinical populations and other endpoints. We have started a collaboration with the Institut Pasteur (GHFC team, T Bourgeron, R Delorme) and the University of Montreal (PPSP team, G Dumas), to characterize differences between normally developing children and children diagnosed with autism spectrum disorders. A wide array of EEG tasks will be used and endpoints (i.e. developmental timepoints) will go beyond the usually accurate diagnosis, focusing on symptom severity and social developmental scores. With the anesthesiology department at the Lariboisière hospital (A Mebazaa, E Gayat, F Vallée) and the cognitive neurology unit (C Paquet) we aim at developing EEG-based models of cognitive decline and dysfunction in two different settings. Postoperative cognitive decline is an important complication after general anesthesia and its antecedents must be better understood. As this might be an indicator for a latent neurodegenerative condition, we plan to use our EEG-based models of both Alzheimer's Disease and Lewy body dementia in which disease progression is an important change over time.

This widening scope calls for a more general methodology as compared to our previous work on DoC. For example, in these conditions involving neurodegenerative problems, we have observed that both subtle and condition-specific spatial patterns matter more than strong and global amplitude changes. To approach these challenges we will draw on our latest M/EEG-methods that were recently developed for population-level modeling of brain health and brain aging [109]. We found that frequency band-specific spatial patterns of M/EEG power spectra conveyed important information of cognitive function (memory and cognitive performance) that were not explained by MRI or fMRI. This was implemented by predicting from a filter-bank of frequency-band-specific source power and source connectivity features. Core challenges to enable clinical translation include lower SNR and absence of individual anatomical MRI scans needed for gold-standard source modeling. Through theoretical analysis, simulations and benchmarks we found [166, 165] that, in M/EEG sensor space, covariance matrices in combination with spatial filtering techniques and Riemannian embeddings provide good workarounds for absent anatomical MRI scans. This covariance-based approach allows to capture fine-grained spatial information related to power and connectivity without performing biophysics-based source localization. Moreover, Riemannian embeddings make predictive modeling from M/EEG covariance matrices more robust to noise, whereas their interpretability is more challenging than that of spatial filters, indicating a direction for further research. Another challenge is given by the limited numbers of labeled samples for supervised learning and EEG-devices with small channel numbers, such as monitoring or user-grade EEG with 2-4 electrodes for which random loss of electrodes can be frequent. In this context, we expect important enhancements from self-supervised learning approaches [75] and deep learning methods for data-augmentation for which we have obtained the first results on non-clinical data. In these settings, the previous elements from classical approaches such as Riemannian geometry or spatial filtering can be readily implemented alongside more involved computations and transformations.

4.3.2 MRI-based modeling of clinical endpoints

Image based biomarkers can be objectively measured and are a sign of normal or abnormal processes, of a condition or disease. Incorporating new potential imaging biomarkers requires several steps, often in parallel and complementary to each other, to be undertaken for translation into clinical practice. These can be divided into the following phases after identification: Development and evaluation, validation, implementation, qualification, and utilization. Our team aims to cross two main translational gaps, that is, the translation from patients first and then to practice. Our aim through our current and active projects is to ensure that potential biomarkers, like the clear delineation of subterritories of the subthalamic nucleus (STN) in pharmaco-resistant Parkinson's disease (PD) patients (i.e.candidates for implantation of a deep brain stimulator) are 'fit for purpose' and associated with the clinical endpoint of interest with the overarching goal being to demonstrated efficacy and health impact. This process is key to the translation into clinical practice

and widespread utilization.

Through the ANR VLFMRI grant we aim to derive new MR imaging-based biomarkers related to prematurity and abnormal neurodevelopment of hospitalized neonates at low magnetic field (20 mTesla). In this setup, the objective is to perform an almost continuous monitoring to detect early signs of adverse events including ischemic stroke or encephalopathy (collaboration with Prof. V. Biran, APHP Robert Debré Hospital). An additional collaboration is already underway with the AP-HP Henri Mondor Hospital (neuroradiologist Dr B. Bapst, doing part of her PhD at NeuroSpin), to achieve high-resolution susceptibility weighted imaging (600 µisotropic) in a scan time of 2m30s for an accurate delineation of the STN in PD patients prior to surgical planning. A database of 123 patients has already been collected using both the standard SWI imaging protocol and ours based on the SPARKLING technology. This annotated database will be key to compare the diagnosis power of our solution with that of the current care, analyse to what extent a higher image resolution is instrumental in providing a more accurate clinical diagnostic, and finally make our protocol more widely accepted in the clinical practice.

Our key contribution in these projects is to translate to the clinical realm both the SPARKLING technology on the acquisition side [130, 86] as well as our PySAP software [111] for MR image reconstruction. In this regard, the recently accepted CEA postdoc funding should help us move the technology to clinical 7T MR Systems (Magnetom Terra Siemens-Healthineers) in the University hospital of Poitiers through a nascent collaboration with Prof. Rémy Guillevin. Their interest is to use the high-resolution SPARKLING SWI protocol at 7T to better delineate the anomalies along the central vein for the diagnostic of multiple sclerosis as the number of anomalies predicts the grade/severity of this inflammatory pathology. On a longer perspective, we aim to generalize the use of our recently DL networks for MR image reconstruction [158, 159] to multiple acquisition setups and other downstream tasks (e.g. motion correction and correction of off-resonance artifacts related to B_0 inhomogeneities).

4.4 From brain images and bio-signals to quantitative biology and physics

Thanks to the developments in subsection 3.1 and subsection 3.4 we aim to approximate more accurately the biophysical models underlying MRI and electrophysiological signals. By estimating quantities grounded in the physics of the data (time, spatial localization, tissue properties) we ambition to offer more actionable outputs for cognitive, clinical and pharmacological applications.

Technologies like 4D SPARKLING should in the future allow us to carry out both fast high resolution multi-parametric quantitative imaging (e.g. T1, T2 and proton density mapping) and laminar (i.e. layer-based) functional imaging in BOLD-fMRI. First, in the mqMRI and fMRI setting, the fourth dimension is respectively the weighting contrast and time axis. mqMRI imaging enables a precise quantification of biomarkers such as iron stores in the pathological brain. Measuring these parameters intra-cortically in Parkinsonian patients defines one of the key challenges in the coming years, especially at 7 Tesla, to earlier stratify the PD patients and the evolution of their disease. Second, a particular attention will be paid to the impact of the developments performed in subsection 3.1 on the statistical sensitivity of brain activity detection, which eventually defines the final validation metric of the data acquisition/image reconstruction pipeline. For this purpose, robust experimental activation protocols such as retinotopic mapping will be used for validation on the 7T scanner and eventually on the 11.7T Iseult MR system. The finest target resolution is 500 μm isotropic in 3D.

Novel development on bi-level optimization for hyper-parameter selection from subsubsection 3.1.2 will bring state-of-the-art inverse methods to end users currently facing the difficulty of performing model selection on empirical data efficiently. This will lead to more accurate quantitative assessments, in sub-millimeters and milliseconds, of where neural activity occurs.

The line of work on inverse problems should also impact how non-invasive neuroimaging and electrophysiology, based on MRI, EEG and MEG, is considered by more traditional neurophysiologists working with animal data. By considering biophysical models of the data and aiming to estimate their parameters from empirical recordings our hope is to present estimates of physical quantities (tissue properties, neural interactions strengths, etc.). The line of work based on stochastic simulation based inference (SBI) can revolutionize the way MEG, EEG and MRI data are apprehended.

For this line of work we will explore the inversion of the models as offered by major software such as The Virtal Brain (TVB) [167] or the Human Neocortical Neurosolver (HNN) [145]. A student from the group of Prof. S. Jones at the origin of the HNN software visited the team in 2022.

5 Social and environmental responsibility

The MIND team has not yet implemented specific guidelines for measuring carbon emission related to its research activities. Team members maximize the use of train for travelling across Europe and try to minimize the number of oversea flights per individual.

6 Highlights of the year

Currently a Research Director within the MIND team at the Inria Saclay center, Demian Wassermann was appointed Deputy Research Director of the DATAIA Institute Paris-Saclay on September 19, 2024.

6.1 Awards

In July 2024, Philippe Ciuciu was awarded one of the two first high-risk structured research projects within the 'Audace!' program of the CEA. His project, titled BrainSync, aims to understand the learning and decision-making mechanisms in the human brain while also developing new rehabilitation protocols for patients suffering from motor disabilities following a stroke, using innovative AI and neuroprosthetics. The project is funded for approximately 4.5 years with a budget of around 5 million euros and involves nine academic and clinical partners, including Inria and CEA, the two public research institutions supporting MIND.

7 New software, platforms, open data

7.1 New software

7.1.1 MNE

Name: MNE-Python

Keywords: Neurosciences, EEG, MEG, Signal processing, Machine learning

Functional Description: Open-source Python software for exploring, visualizing, and analyzing human neurophysiological data: MEG, EEG, sEEG, ECoG, and more.

Release Contributions: https://mne.tools/stable/whats_new.html

URL: https://mne.tools/

Contact: Alexandre Gramfort

Partners: HARVARD Medical School, New York University, University of Washington, CEA, Aalto university, Telecom Paris, Boston University, UC Berkeley, Macquarie University, University of Oregon, Aarhus University

7.1.2 NeuroLang

Name: NeuroLang

Keywords: Neurosciences, Probabilistic Programming, Logic programming

Functional Description: NeuroLang is a probabilistic logic programming system specialised in the analysis of neuroimaging data, but not exclusively determined by it.

Release Contributions: https://neurolang.github.io/

URL: https://neurolang.github.io/#

Contact: Demian Wassermann

7.1.3 Nilearn

Name: NeuroImaging with scikit learn

Keywords: Health, Neuroimaging, Medical imaging

Functional Description: NiLearn is the neuroimaging library that adapts the concepts and tools of scikit-learn to neuroimaging problems. As a pure Python library, it depends on scikit-learn and nibabel, the main Python library for neuroimaging I/O. It is an open-source project, available under BSD license. The two key components of NiLearn are i) the analysis of functional connectivity (spatial decompositions and covariance learning) and ii) the most common tools for multivariate pattern analysis. A great deal of efforts has been put on the efficiency of the procedures both in terms of memory cost and computation time.

Release Contributions: HIGHLIGHTS - Updated docs with a new theme using furo. - permuted_ols and non_parametric_inference now support TFCE statistic. - permuted_ols and non_parametric_inference now support cluster-level Family-wise error correction. - save_glm_to_bids has been added, which writes model outputs to disk according to BIDS convention.

NEW - save_glm_to_bids has been added, which writes model outputs to disk according to BIDS convention. - permuted_ols and non_parametric_inference now support TFCE statistic. - permuted_ols and non_parametric_inference now support cluster-level Family-wise error correction. - Updated docs with a new theme using furo.

See all details in https://nilearn.github.io/stable/changes/whats_new.html

URL: http://nilearn.github.io/

Contact: Bertrand Thirion

Participants: Alexandre Abraham, Alexandre Gramfort, Bertrand Thirion, Elvis Dohmatob, Fabian Pedregosa Izquierdo, Gael Varoquaux, Loic Esteve, Michael Eickenberg, Virgile Fritsch

7.1.4 Benchopt

Keywords: Benchmarking, Machine learning, Optimization

Functional Description: BenchOpt is a package to simplify, make more transparent and more reproducible the comparisons of optimization algorithms. It is written in Python but it is available with many programming languages. So far it has been tested with Python, R, Julia and compiled binaries written in C/C++ available via a terminal command. If it can be installed via conda, it should just work!

BenchOpt is used through a simple command line and ultimately running and replicating an optimization benchmark should be as easy a cloning a repo and launching the computation with a single command line. For now, BenchOpt features benchmarks for around 10 convex optimization problems and we are working on expanding this to feature more complex optimization problems. We are also developing a website to display the benchmark results easily.

Release Contributions: https://github.com/benchopt/benchopt/releases/tag/1.5.1

News of the Year: In June 2024, we organized a benchopt benchmarking sprint with around 40 participants.

During this sprint, we developed new benchmarks, gathered feedback, and saw which benchmarking tools were most needed. Most of the participants had never worked with benchopt. They aimed to create new benchmarks using benchopt, focusing on ML tasks such as time-series forecasting or tabular learning. The participants were all able to produce benchmarks within a few hours! Note that an important aspect of this benchmarking initiative is that the goal is not to re-code the various methods but to aggregate them from the different sources from which they are available.

Take aways from this sprint have been discussed in a blog post accessible here: https://notes.inria.fr/_fA4TnmHScS

Publication: hal-03830604

Contact: Thomas Moreau

Participants: Thomas Moreau, Alexandre Gramfort, Mathurin Massias, Badr Moufad

7.1.5 Scikit-learn

Keywords: Clustering, Classification, Regression, Machine learning

Scientific Description: Scikit-learn is a Python module integrating classic machine learning algorithms in the tightly-knit scientific Python world. It aims to provide simple and efficient solutions to learning problems, accessible to everybody and reusable in various contexts: machine-learning as a versatile tool for science and engineering.

Functional Description: Scikit-learn can be used as a middleware for prediction tasks. For example, many web startups adapt Scikitlearn to predict buying behavior of users, provide product recommendations, detect trends or abusive behavior (fraud, spam). Scikit-learn is used to extract the structure of complex data (text, images) and classify such data with techniques relevant to the state of the art.

Easy to use, efficient and accessible to non datascience experts, Scikit-learn is an increasingly popular machine learning library in Python. In a data exploration step, the user can enter a few lines on an interactive (but non-graphical) interface and immediately sees the results of his request. Scikitlearn is a prediction engine . Scikit-learn is developed in open source, and available under the BSD license.

URL: `http://scikit-learn.org`

Publications: hal-00650905, hal-00856511, hal-01093971

Contact: Gael Varoquaux

Participants: Thomas Moreau, Jerome Dockes, Alexandre Gramfort, Bertrand Thirion, Gael Varoquaux, Loic Esteve, Olivier Grisel, Guillaume Lemaitre, Jeremie Du Boisberranger, Julien Jerphanion

Partners: Axa, BNP Parisbas Cardif, Dataiku, Nvidia, Chanel, Probabl

7.1.6 joblib

Keywords: Parallel computing, Cache

Functional Description: Facilitate parallel computing and caching in Python.

URL: `https://joblib.readthedocs.io/en/latest/`

Contact: Thomas Moreau

Participant: Thomas Moreau

Partner: Probabl

7.1.7 MRI-NUFFT

Keywords: Brain MRI, NUFFT, Trajectory Generation

Functional Description: MRI-NUFFT is a python package that extends various NUFFT (Non-Uniform Fast Fourier Transform) python bindings used for MRI reconstruction. It provides a unified interface with a large number of backends with implementations ranging from CPU to GPU.

In particular, it provides a unified interface for all the methods, with extra features such as coil sensitivity, density compensated adjoint and off-resonance corrections (for static B0 inhomogeneities). Additionally, useful IO tools like reading a k-space sampling trajectory and writing a binary file for run on MR scanner is also offered. Finally, it helps algorithmically speed up MR image reconstruction algorithms through fast ways to estimate preconditioning weights, also known as density compensators for a given sampling pattern.

Release Contributions: MRI-NUFFT now provides a physical model of the MRI acquisition processes, including multi-coil acquisition and static-field inhomogeneities. This model is compatible with the NUFFT libraries, and can be used to simulate the acquisition of MRI data, or to reconstruct data from a given set of measurements. MRI-NUFFT comes with a wide variety of non-Cartesian trajectory generation routines that have been gathered from the literature. It also provides ways to extend existing trajectories and export them to specific formats, for use in other toolkits and on MRI hardware. Finally, MRI-NUFFT provides automatic differentiation for all NUFFT backends, with respect to both gradients and data (image or k-space). This enables efficient backpropagation through NUFFT operators and supports research on learned sampling model and image reconstruction network.

URL: https://mind-inria.github.io/mri-nufft/

Contact: Chaithya Giliyar Radhkrishna

7.1.8 SPARKLING

Name: Spreading Projection Algorithm for Rapid K-space sampLING

Keywords: Brain MRI, MRI, Optimization

Scientific Description: This python package allows us to generate "SPARKLING" curves as a new type of non-Cartesian trajectories to perform a more efficient sampling in 2D and 3D for anatomical imaging while using the same number of samples for a limited time budget. These segmented curves are obtained using a projection method on measure sets which offers three main advantages: i) generating segmented Non-Cartesian trajectories along a chosen density, ii) meeting the hardware constraints on the magnetic field gradients (magnitude, slew rate), iii) performing a fast coverage of k-space.

Functional Description: This python package implements "SPARKLING": an optimization driven method to obtain hardware compliant sampling curves that globally satisfy a user specified target sampling density. The resulting non-cartesian sampling curves can be used to efficiently undersample and speed up acquisitions on an MR scanner. This method is generic enough that it can be applied to any of the imaging modalities in MR.

Publications: hal-01577200v1, hal-02067080v1, hal-03090471v2, hal-04370475v1, hal-02361265v1

Contact: Chaithya Giliyar Radhkrishna

7.1.9 PySAP

Name: Python Sparse data Analysis Package

Keywords: Image reconstruction, Image compression

Functional Description: The PySAP (Python Sparse data Analysis Package, https://github.com/CEA-COSMIC/pysap) open-source image processing software package has been developed for the 3 years between the Compressed Sensing group at Iniria-CEA Parietal team led by Philippe Ciuciu and the CosmoStat team (CEA/IRFU) led by Jean-Luc Statck. It has been developed for the COmpressed Sensing for Magnetic resonance Imaging and Cosmology (COSMIC) project. This package provides a set of flexible tools that can be applied to a variety of compressed sensing and image reconstruction problems in various research domains. In particular, PySAP offers fast wavelet transforms and a range of integrated optimisation algorithms. It also offers a variety of plugins for specific application domains: on top of Pysap-MRI and PySAP-astro plugins, several complementary modules are now in development for electron tomography and electron microscopy for CEA colleagues. In October 2019, PySAP has been released on PyPi (https://pypi.org/project/python-pySAP/, currently version 0.0.3) and in conda (https://anaconda.org/agrigis/python-pysap).

The Pysap-MRI has been advertised through a specific abstract accepted to the next workshop of ISMRM on Data Sampling & Image Reconstruction in late January 2020. It will be presented during a power pitch session together wih an hands-on demo session using JuPyter notebooks.

Contact: Philippe Ciuciu

Partner: CEA

7.1.10 SNAKE

Name: Simulator from neuro-activation to K-space Exploration

Keywords: FMRI, NUFFT

Functional Description: We propose a new, modular, open-source, Python-based 3D+time fMRI data simulation software, SNAKE-fMRI, which stands for Simulator from Neurovascular coupling to Acquisition of K-space data for Exploration of fMRI acquisition techniques. Unlike existing tools, the goal here is to simulate the complete chain of fMRI data acquisition, from the spatio-temporal design of evoked brain responses to various multi-coil k-space data 3D sampling strategies, with the possibility of extending the forward acquisition model to various noise and artifact sources while remaining memory-efficient. By using this in silico setup, we are thus able to provide realistic and reproducible ground truth for fMRI reconstruction methods in 3D accelerated acquisition settings and explore the influence of critical parameters, such as the acceleration factor and signal-to-noise ratio (SNR), on downstream tasks of image reconstruction and statistical analysis of evoked brain activity. We present three scenarios of increasing complexity to showcase the flexibility, versatility, and fidelity of SNAKE-fMRI: From a temporally-fixed full 3D Cartesian to various 3D non-Cartesian sampling patterns, we can compare — with reproducibility guarantees — how experimental paradigms, acquisition strategies and reconstruction methods contribute and interact together, affecting the downstream statistical analysis.

URL: https://paquiteau.github.io/snake-fmri/

Contact: Philippe Ciuciu

7.2 Open data

Participants: Bertrand Thirion, Ana Fernanda Ponce, Himanshu Aggarwal.

Main External Collaborators: Ana-Luisa Pinho *(Western University, Canada)*, Lucie Hertz-Pannier *(CEA/NeuroSpin)*.

The MIND is an an open data provider as it produced the Individual Brain Charting (IBC) data set over the last year in the context of the Human Brain Project. The IBC data set consists of anatomical and functional brain MR images collected on twelve healthy volunteers. It is quite unique as approxilmately 50 to 60 different behaviarol protocols have been run on these individuals. The IBC data set can be uploaded here.

8 New results

8.1 Bridging Cartesian and non-Cartesian sampling in MRI

Participants: Chaithya Giliyar Radhakrishna, Philippe Ciuciu.

Main External Collaborators: Alexandre Vignaud *(CEA/NeuroSpin)*, Aurélien Massire *(Siemens-Healthineers, France)*.

MRI is a widely used neuroimaging technique used to probe brain tissues, their structure and provide diagnostic insights on the functional organization as well as the layout of brain vessels. However, MRI relies on an inherently slow imaging process. Reducing acquisition time has been a major challenge in high-resolution MRI and has been successfully addressed by Compressed Sensing (CS) theory. However, most of the Fourier encoding schemes under-sample existing k-space trajectories which unfortunately will never adequately encode all the information necessary. Recently, the MIND team has addressed this crucial issue by proposing the Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) for 2D/3D non-Cartesian T2* and susceptibility weighted imaging (SWI) at 3 and 7 Tesla (T) [129, 130], [6].

These advancements have interesting applications in cognitive and clinical neuroscience. However, the original SPARKLING trajectories are prone to off-resonance effects due to susceptibility artifacts. Therefore, in 2023, two years ago, we extended the original SPARKLING methodology along two diretions : and developed the MORE-SPARKLING (Minimized Off-Resonance Effects) approach to correct for these artifacts. The fondamental idea implemented in MORE-SPARKLING is to make different k-space trajectories more homogeneous in time, in the sense that samples supported by different trajectories that are close in k-space must be collected at approximately the same time point. This allows us to mitigate the issue of off-resonance effects (signal void, geometric distortions) without increasing the scan time, as MORE-SPARKLING trajectories have exactly the same duration as their SPARKLING ancestor. This approach was published [5]. Additionally, we also extended SPARKLING in another direction, namely the way we sample the center of k-space and proposed the GoLF-SPARKLING version in the same paper [5]. The core idea was to reduce the over-sampling of the center of k-space and grid it to collect Cartesian data and make notably the estimation of sensitivity maps in multicoil acquisition easier.

In 2024, we have merged GoLF-SPARKLING with well-established clinically used parallel imaging techniques like GRAPPA and CAIPIRINHA. The latter are mostly limited to Cartesian sampling, and their extension to non-Cartesian sampling is not direct.In [52] we proposed to extend the SPARKLING frameworkto ensure GRAPPA accelerated Cartesian sampling in the k-space center while allowing non-Cartesian sampling in the periphery. We can now enforce Cartesian acceleration in the center of k-space through affine constraints. Using 2×2 GRAPPA with SPARKLING trajectories for anatomical MRI with MPRAGE sequences enables 10x acceleration, achieving 1mm isotropic whole-brain scans in 1 minute (see Fig. 1).

8.2 SPARKLING for fMRI

Participants: Zaineb Amor, Philippe Ciuciu.

Main External Collaborators: Alexandre Vignaud *(CEA/NeuroSpin)*.

Additionally, we have shown that 3D-SPARKLING is a viable imaging technique and good alternative to Echo Planar Imaging for resting-state and task-based fMRI [11] and [12]. This is illustrated in Fig. 2 during a retinotopic mapping experiment which consists in mapping the retina to the primary visual cortex. These results have been obtained at a 1mm isotropic resolution both for EPI and SPARKLING acquisitions.

8.3 Deep Learning for 3D Non-Cartesian MR Image reconstruction

Participants: Asma Tanabene, Chaithya Giliyar Radhakrishna, Philippe Ciuciu.

External Collaborators: Aurélien Massire *(Siemens-Healthineers, France)*.

Deep learning and notably unrolled neural netowrk architectures have shown great promise for MRI reconstruction from undersampled data. However, there is a lack of research on validating their performance in the 3D parallel imaging acquisitions with non-Cartesian undersampling. In addition, the artifacts and the resulting image quality depend on the under-sampling pattern. To address this uncharted territory, in 2024 we extended the Non-Cartesian Primal-Dual Network (NC-PDNet) [160], to a 3D multi-coil acquisition setting. We evaluated the impact of channel-specific versus channel-agnostic training configurations and examined the effect of coil compression. Finally, using the publicly available Calgary-Campinas dataset, we benchmarked four distinct non-Cartesian undersampling patterns, with an acceleration factor of six. Our results in Fig. 3 showed that NC-PDNet trained on compressed data with varying input channel numbers achieves an average PSNR of 42.98dB for 1 mm isotropic 3- channel whole-brain 3D reconstruction. With an inference time of 4.95sec and a GPU memory usage of 5.49 GB, our approach demonstrates significant potential for clinical research application.

8.4 Fast reconstructions of ultra-high resolution MR data from the 11.7T Iseult scanner

Participants: Chaithya Giliyar Radhakrishna, Philippe Ciuciu.

External Collaborators: Alexandre Vignaud *(CEA/NeuroSpin)*, Franck Mauconduit *(CEA/NeuroSpin)*.

Open-source MR reconstruction tools often fail to efficiently utilize GPU resources and lack support for generalized GRAPPA implementations. Many tools are limited to 2D or 3D reconstruction, and few incorporate advanced techniques such as 2D-CAIPIRINHA, which enhances imaging capabilities. Our new open source gGRAPPA GPU accelerated Python package aims

to provide a fast, flexible, open-source tool for generalized GRAPPA/CAIPI reconstruction [45]. Using PyTorch, gGRAPPA runs multiple convolutional windows in batch mode to optimize GPU memory usage and accelerate reconstruction times. gGRAPPA achieves up to a 65x speedup over CPU implementations and a 6x speedup compared to non-batched GPU methods, enabling efficient and fast reconstruction of MRI scans. This tool has allowed us to outperform the Siemens image reconstructor, in terms of speed, on the world premiere 11.7T MR system (Iseult scanner available at NeuroSpin) to reconstruct the first in vivo T_2^* weighted MR images at a 200-µm in plane resolution.

8.5 SNAKE-fMRI: A realistic fMRI data simulator for high resolution functional imaging

Participants: Pierre-Antoine Comby, Philippe Ciuciu.

External Collaborators: Alexandre Vignaud *(CEA/NeuroSpin)*.

Functional Magnetic Resonance Imaging (fMRI) has emerged as a powerful non-invasive tool in neuroscience, enabling neuroscientists to understand human brain functions. However, fMRI acquisition and image reconstruction techniques are complex to optimize and benchmark due to the lack of ground truth that produces absolute and quantitative metrics. Repeating in-vivo experiments may face the issue of limited reproducibility, which is time-consuming and expensive. fMRI simulators have been developed to generate synthetic fMRI images, where brain responses are artificially added to existing or artificial data. However, they don't simulate the complete MR acquisition process, lack flexibility, integration with post-processing, and computational efficiency. Exploring new acceleration schemes in the acquisition setting and innovative reconstruction methods with these tools is not feasible. To address these unmet needs, in 20204 we have developed SNAKE-fMRI [51], an open-source fMRI simulator that operates in both image and k-space domains to yield realistic synthetic fMRI data. Its flexibility allows us to investigate various acquisition setups regarding SNR and acceleration factors to reach higher spatial and temporal resolution and validate reconstruction methods against those scenarios. Its key principles are summarized in Fig. 5 and some comparative results are shown in Fig. 6.

The interest of this simulator will be to provide ground truth data to train deep learning models for fMRI reconstruction in 2025.

8.6 NeuroConText: Contrastive text-to-brain mapping for neuroscientific literature

Participants: Bertrand Thirion, Demian Wassermann, Fatemeh Ghayem, Raphael Meudec.

Hundreds of neuroscientific articles are published annually, highlighting the continuous growth and expansion of knowledge in this field. These contributions from scientists and researchers provide new insights and findings that enhance our understanding of brain functions. Meta-analysis is a statistical tool that combines results from multiple studies to improve the reliability and generalizability of neuroscientific findings. It serves three key purposes: First, it synthesizes information from the literature to build the state of the art by offering a consolidated view of what is currently known. This allows researchers to see where the field stands collectively and highlights consensus or inconsistencies. Second, meta-analysis provides context to interpret new results by comparing novel experimental data with existing patterns in the meta-analysis of the literature,

which clarifies how new data align with or deviate from established findings. Third, meta-analysis generates hypotheses on candidate brain regions or relevant cognitive domains by revealing patterns that might be central to a phenomenon.

In [34], we introduced NeuroConText, a novel coordinate-based meta-analysis (CBMA) tool to bridge the three heterogeneous modalities commonly found in neuroscientific studies: text, reported brain activation coordinates, and brain images. NeuroConText uses neuroscientific articles to extract their text and activation coordinates. We benefited from the embeddings of advanced large language models (LLM) for text feature representation. Additionally, we used Kernel Density Estimation (KDE) to reconstruct brain maps from the coordinates [169, 168]. To address the high dimensionality of the brain images reconstructed by KDE, we employed the dictionary of functional modes (DiFuMo), a probabilistic atlas that effectively reduces data dimensionality [98].

Then, NeuroConText defines a shared latent space between text and coordinates, using contrastive learning to retrieve the brain activation coordinates corresponding to the input text. NeuroConText can analyze long texts, leveraging the complete information in articles' text to enhance the accuracy of text-to-brain associations. By incorporating advanced language models like Mistral-7B, it processes complex neuroscientific text and extracts the semantic in the text [124]. NeuroConText considerably outperforms existing regression-based state-of-the-art methods NeuroQuery and Text2Brain in associating text with brain activations, achieving a threefold improvement in the retrieval task. To improve our understanding of the internal mechanisms of the NeuroConText model, we also evaluated its ability to reconstruct brain activation contrasts from text latent representations using descriptions of the NeuroVault dataset [114]. The quality of these reconstructed activation maps is found to be comparable with state-of-the-art baselines.

8.7 Improved priors for inverse problem resolution

Participants: Matthieu Terris, Thomas Moreau.

External Collaborators: Tachella Julian *(CNRS)*, Pustelnik Nelly *(CNRS)*, Ulugbek Kamilov *(Meta)*.

Selecting an appropriate prior to compensate for information loss due to the measurement operator is a fundamental challenge in inverse problems. Implicit priors based on denoising neural networks have become central to widely-used frameworks such as Plug-and-Play (PnP) algorithms. During this year, we made several progress to better design implicit priors based on existing denoisers. First, we proposed to enforce equivariance to certain groups of transformations (rotations, reflections, and/or translations) on the denoiser used as implicit priors. This simple procedure strongly improves the stability of the algorithm as well as its reconstruction quality. We derived theoretical insights which highlight the role of equivariance on better performance and stability, and experiments on multiple imaging modalities and denoising networks show numerically these benefits. Then, we introduce Fixed-points of Restoration (FiRe) priors as a new framework for expanding the notion of priors in PnP to general restoration models beyond traditional denoising models. The key insight behind FiRe is that natural images emerge as fixed points of the composition of a degradation operator with the corresponding restoration model. Adopting this fixed-point perspective, we show how various restoration networks can effectively serve as priors for solving inverse problems. Experimental results validate the effectiveness of FiRe across various inverse problems, establishing a new paradigm for incorporating pretrained restoration models into PnP-like algorithms.

8.8 Statistically Valid Variable Importance Assessment through Conditional Permutations

Participants: Bertrand Thirion, Ahmad Chamma.

External Collaborators: Denis Engemann *(Roche)*.

Variable importance assessment has become a crucial step in machine-learning applications when using complex learners, such as deep neural networks, on large-scale data. Removal-based importance assessment is currently the reference approach, particularly when statistical guarantees are sought to justify variable inclusion. It is often implemented with variable permutation schemes. On the flip side, these approaches risk misidentifying unimportant variables as important in the presence of correlations among covariates. Here we develop a systematic approach for studying Conditional Permutation Importance (CPI) that is model agnostic and computationally lean, as well as reusable benchmarks of state-of-the-art variable importance estimators. We show theoretically and empirically that CPI overcomes the limitations of standard permutation importance by providing accurate type-I error control. When used with a deep neural network, CPI consistently showed top accuracy across benchmarks. An experiment on real-world data analysis in a largescale medical dataset showed that CPI provides a more parsimonious selection of statistically significant variables. Our results suggest that CPI can be readily used as drop-in replacement for permutation-based methods.

8.9 False Discovery Proportion control for aggregated Knockoffs

Participants: Bertrand Thirion, Alexandre Blain.

Main External Collaborators: Pierre Neuvial *(IMT, Univ. Toulouse)*, Olivier Grisel *(Inria, Soda team)*.

Controlled variable selection is an important analytical step in various scientific fields, such as brain imaging or genomics. In these high-dimensional data settings, considering too many variables leads to poor models and high costs, hence the need for statistical guarantees on false positives. Knockoffs are a popular statistical tool for conditional variable selection in high dimension. However, they control for the expected proportion of false discoveries (FDR) and not their actual proportion (FDP). We present a new method, KOPI, that controls the proportion of false discoveries for Knockoff-based inference. The proposed method also relies on a new type of aggregation to address the undesirable randomness associated with classical Knockoff inference. We demonstrate FDP control and substantial power gains over existing Knockoff-based methods in various simulation settings and achieve good sensitivity/specificity tradeoffs on brain imaging and genomic data.

8.10 Clinical biomarkers for epilepsy

Participants: Sheng H Wang, Philippe Ciuciu.

External Collaborators: Matias Palva *(Aalto University, Finland)*, Satu Palva *(Neuroscience Center, Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Finland)*.

Postsurgical seizure freedom in drug-resistant epilepsy (DRE) patients varies from 30% to 80%, implying that in many cases the current approaches fail to fully map the epileptogenic zone (EZ). We aimed to advance a novel approach to better characterize epileptogenicity and investigate whether the EZ encompasses a broader epileptogenic network (EpiNet) beyond the seizure zone (SZ) that exhibits seizure activity. We first used computational modeling to test putative complex systems-driven and systems neuroscience-driven mechanistic biomarkers for epileptogenicity. We then used these biomarkers to extract features from resting-state stereo-electroencephalograms recorded from DRE patients and trained supervised classifiers to localize the SZ against gold standard clinical localization. To further explore the prevalence of pathological features in an extended brain network outside of the clinically identified SZ, we also used unsupervised classification. Supervised SZ classification trained on individual features achieved accuracies of 0.6–.07 area under the receiver operating characteristic curve (AUC).* Combining all criticality and synchrony features further improved the AUC to 0.85. Unsupervised classification discovered an EpiNet-like cluster of brain regions, in which 51% of brain regions were outside of the SZ. Brain regions in the EpiNet-like cluster engaged in interareal hypersynchrony and locally exhibited high-amplitude bistability and excessive inhibition, which was strikingly similar to the high seizure risk regime revealed by our computational modeling. The finding that combining biomarkers improves SZ localization accuracy indicates that the novel mechanistic biomarkers for epileptogenicity employed here yield synergistic information. On the other hand, the discovery of SZ-like brain dynamics outside of the clinically defined SZ provides empirical evidence of an extended pathophysiological EpiNet.

9 Bilateral contracts and grants with industry

Participants: Philippe Ciuciu, Asma Tanabene, Thomas Moreau.

9.1 Bilateral contracts with industry

9.1.1 Siemens Healthineers & AI lab (Princeton, USA)

Since Fall 2019, Philippe Ciuciu has actively collaborated with the Siemens-Healthineers AI lab, led by Mariappan Nadar. In this context, a new CIFRE PhD student, Mrs Asma Tanabene, has joint the MIND team as a CIFRE PhD student under their joint supervision to work on 3D sclalable deep learning architecture for high-resolution multicoil MR image reconstruction at 3 Tesla and beyond. On top of the PhD funding, this contract has generated 50k€ for Mind, a grant that is hosted at CEA/NeuroSpin.

9.1.2 Saint Gobain Research (SGR)

There is currently a consulting contract between SGR and Mind (Thomas Moreau) to provide an expertise in machine learning to process temporal data, numerical optimization and scientific computing. The expertise is provided one half-day per month, in SGR offices, and it consists in scientific discussion sessions on the ML projects leaded by SGR data scientists.

9.1.3 Apple

A research collaboration between Apple (Pierre Ablin) and Mind (Thomas Moreau) has been established. The goal is to develop research ideas on bilevel optimization, in particular for the problem of data curation for the training of large models with massive and heterogeneous datasets. This collaboration is being funded by Apple (150k€) and allowed to hire a post-doc (Baptiste Goujaud).

9.2 Bilateral Grants with Industry

9.2.1 Google

MIND (Thomas Moreau) received a 30k€ donation from Google to support its open source activity around `benchopt`. In particular, the grant aims to support the organization of coding and benchmarking sprints around benchopt, the development of visualization tools, and the benchmarking of bilevel solvers, in particular the ones using `jaxopt`.

10 Partnerships and cooperations

10.1 International initiatives

10.1.1 Inria associate team not involved in an IIL or an international program

NeuroMind

Title: Precision mapping of the Brain by Neuromod & Mind

Duration: 2022 ->

Coordinator: Pierre BELLEC (pierre.bellec@criugm.qc.ca)

Partners:

- Université de Montréal Montréal (Canada)

Inria contact: Bertrand Thirion

Summary: Among the main advances of the last decade, the development of powerful AI systems for vision, language processing, as well as reinforcement learning, have led to sophisticated cognitive systems that can be compared to the human brain, and sometimes surpass human performance. Brain/AI system comparison is a great opportunity for AI and for neuroscience. One of the most urgent tasks for cognitive neuroscience is thus to put together datasets that probe the brain system and are comprehensive enough to allow a reliable comparison of brain activity to the representations generated by AI systems. To address this endeavor, Parietal and NeuroMod have launched ambitious data acquisitions initiative (individual Brain CHarting and Courtois Neuromod), that consist in collecting huge amounts of brain data in few participants. These unprecedented data collection efforts bring novel challenges for data analysis: handling TB-scale data, automation, and better integration of analysis pipeline. Software such as Nilearn and MNE increasingly face the challenge of scaling up to larger datasets. Addressing this challenge in the context of IBC and Courtois Neuromod is thus a unique opportunity.

10.2 International research visitors

10.2.1 Visits of international scientists

Other international visits to the team

[•] David Degras

Status Associate Professor in the Department of Mathematics

Institution of origin: University of Massachusetts Boston

Country: USA

Dates: from 01/09/2023 to 30/06/2024

Context of the visit: David is financially supported by the DataIA program and the Inria Saclay Ile-de-France center. His research projects in the team will revolve around machine learning and optimization for computational neuroimaging. For a smooth integration in the team, he is currently involved in dynamic functional connectivity analysis from fMRI data. Next he will be contributing to two ongoing research projects, namely (i) robustness in multifractal analysis of MEG/EEG signals and (ii) functional alignment of brain images. The first project will crucially engage with issues of robustness in learning. The second will largely revolve around optimization (including combinatorial optimization for the choice of barycenter and optimization of hyperparameters). Both projects will investigation the reproducibility of the obtained results as well as out-of-sample validation with independent data. Together, these aspects address the following points of the subject area:

- Automatic learning and hyper-optimization
- Optimization for learning, e.g., improvements in stochastic gradient methods, Bayesian optimization), combinatorial optimization
- Reproducibility and robust learning
- Statistical Inference and Validation.

The proposed projects both contain: (i) an important methodological component of general interest to the statistical and machine learning community, and (ii) an applied neuroimaging component that serves the long-term goal of promoting health (for example, by better understanding the neural foundations of psychological/neurodegenerative pathologies and by better predicting therapeutic/surgical outcomes for people based on their brain function and anatomy).

Mobility program/type of mobility: sabbatical, i.e. invited Professor.

[•] Jean Faber Ferreira de Abreu

Status Associate Professor in the Department of Neurology and Neurosurgery

Institution of origin: University Federal of Sao Paulo (UNIFESP)

Country: Brazil

Dates: from 13/09/2024 to 26/09/2024

Context of the visit: For his stay, Prof. Faber was supported by Inria the Inria Saclay Ile-de-France center. His research projects involve brain computer interfaces, machine learning to decode brain activity from surfacic and intracranial EEG data. In this regards, he is tighlty connected to the BrainSync project led by Philippe Ciuciu. Additionally, as a former visiting scientist at CEA-Clinatec in Grenoble, Prof. Faber has a good understanding and mastery of rehabilitation protocols for motor deprived patients. As the BrainSync projects involve both NeuroSpin and Clinatec, this project aims to foster the collaboration between Prof Faber and French groups to develop new biomarkers based on brain activity that are able to capture motor rehabilitation in Stroke patients. These biomarkers must be computed online in real time from recoprdings collected on implanted WIMAGINE neuroprotheses.

Mobility program/type of mobility: research stay, invited Professor.

10.3 European initiatives

10.3.1 Horizon Europe

EBRAINS 2.0 EBRAINS 2.0 project on cordis.europa.eu

Title: EBRAINS 2.0: A Research Infrastructure to Advance Neuroscience and Brain Health

Duration: From January 1, 2024 to December 31, 2026

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- MEDIZINISCHE UNIVERSITAET WIEN, Austria
- CONVELOP - COOPERATIVE KNOWLEDGE DESIGN GMBH, Austria
- REGION HOVEDSTADEN (REGIONH), Denmark
- HEINRICH-HEINE-UNIVERSITAET DUESSELDORF (UDUS), Germany
- THE CHANCELLOR MASTERS AND SCHOLARS OF THE UNIVERSITY OF CAMBRIDGE, United Kingdom
- INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM), France
- UNIVERSITE LYON 1 CLAUDE BERNARD, France
- KUNGLIGA TEKNISKA HOEGSKOLAN (KTH), Sweden
- THE UNIVERSITY OF MANCHESTER (UNIVERSITY OF MANCHESTER), United Kingdom
- EUROPEAN ACADEMY OF NEUROLOGY, Austria
- FORSCHUNGSZENTRUM JULICH GMBH (FZJ), Germany
- KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN - KNAW (KNAW), Netherlands
- UNIVERSITEIT GENT (UGent), Belgium
- TECHNISCHE UNIVERSITAET MUENCHEN (TUM), Germany
- COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (CEA), France
- EBRAINS (EBRAINS), Belgium
- ASSISTANCE PUBLIQUE HOPITAUX DE PARIS, France
- RUPRECHT-KARLS-UNIVERSITAET HEIDELBERG (UHEI), Germany
- ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS (ATHENA - RESEARCH AND INNOVATION CENTER), Greece
- UNIVERSITAET BERN, Switzerland
- RISE RESEARCH INSTITUTES OF SWEDEN AB (RISE), Sweden
- UNIVERSITA DEGLI STUDI DI TORINO (UNITO), Italy
- UNIVERSITEIT MAASTRICHT, Netherlands
- HITS GGMBH (HITS), Germany
- ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL), Switzerland
- UNIVERSITA DEGLI STUDI DI PADOVA (UNIPD), Italy
- CENTRE HOSPITALIER UNIVERSITAIRE VAUDOIS (CHUV), Switzerland
- NORGES MILJO-OG BIOVITENSKAPLIGE UNIVERSITET (NMBU), Norway
- LIETUVOS SVEIKATOS MOKSLU UNIVERSITETAS (LSMU), Lithuania
- UNIVERSITAT TRIER, Germany
- MEDIZINISCHE UNIVERSITAT INNSBRUCK (MUI), Austria
- TAMPEREEN KORKEAKOULUSAATIO SR (TAMPERE UNIVERSITY), Finland

- UNIVERSITE GRENOBLE ALPES (UGA), France
- CONSIGLIO NAZIONALE DELLE RICERCHE (CNR), Italy
- UNIVERSITE DE BORDEAUX (UBx), France
- UNIVERSITA DEGLI STUDI DI PAVIA (UNIPV), Italy
- FONDAZIONE PER LA RICERCA BIOMEDICA AVANZATA ONLUS (VIMM), Italy
- EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH (ETH Zürich), Switzerland
- CHARITE - UNIVERSITAETSMEDIZIN BERLIN, Germany
- UNIVERSITAETSKLINIKUM HAMBURG-EPPENDORF (UKE), Germany
- UNIVERSITETET I OSLO (UNIVERSITY OF OSLO), Norway
- POLITECNICO DI MILANO (POLIMI), Italy
- UNIVERSITA DEGLI STUDI DI MILANO (UMIL), Italy
- KAROLINSKA INSTITUTET (KI), Sweden
- FONDEN DEMOCRACY X (DEMOCRACY X), Denmark
- KLINIKUM RECHTS DER ISAR DER TECHNISCHEN UNIVERSITAT MUNCHEN (TUM-MED), Germany
- PROTISVALOR MEDITERRANEE SAS (PVM), France
- UNIVERSIDAD REY JUAN CARLOS (URJC), Spain
- UNIVERSITE D'AIX MARSEILLE (AMU), France
- UNIVERSITY COLLEGE LONDON, United Kingdom
- CINECA CONSORZIO INTERUNIVERSITARIO (CINECA), Italy
- UNIVERSITAETSKLINIKUM FREIBURG (UKLFR), Germany
- CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS (CNRS), France
- KATHOLIEKE UNIVERSITEIT LEUVEN (KU Leuven), Belgium
- CODEMART SRL (CODEMART), Romania
- UNIVERSIDAD POLITECNICA DE MADRID (UPM), Spain
- LABORATORIO EUROPEO DI SPETTROSCOPIE NON LINEARI (LENS), Italy
- AZIENDA SANITARIA UNIVERSITARIA FRIULI CENTRALE (AZIENDA SANITARIA UNIVERSITARIA FRIULI CENTRALE), Italy
- TECHNISCHE UNIVERSITAET DRESDEN (TUD), Germany

Inria contact: Bertrand Thirion

Coordinator:

Summary: EBRAINS is a collaborative European Research Infrastructure designed to advance and accelerate progress in neuroscience and brain health. This innovative infrastructure, a legacy of the Human Brain Project (HBP), is an ecosystem where researchers, clinicians and experts from various disciplines converge to explore and analyze brain complexity – from molecular and cellular levels to the functioning of the entire organ. Therefore, the project aims to create a new standard for brain atlases from the micro- to the macro-scale, link foundational multi-level data and connectomes in the healthy and pathological brain with atlases and models, create digital twins through modelling and simulation as well as unique, excellent, and preferred services for FAIR neuroscience data. The overarching goal of EBRAINS 2.0 is to foster a deeper understanding of brain structure and function with dedicated and mature software tools, to facilitate the development of more effective treatments, new drugs, diagnostics and preventive measures for neuro-psychiatric disorders. We expect that EBRAINS 2.0 catalyzes progress in the field of large-scale models running on

HPC towards Exascale and leads to innovative solutions for neuro-inspired computing, and cognitive technologies such as neurorobotics and AI. Sophisticated digital modeling and data analytics capabilities will benefit communities beyond neuroscience, such as biomedicine. We will advance EBRAINS technology, platform services and the base infrastructure roadmap, educate and train a new community of users and developers from academia, industry and SMEs, and ensure knowledge transfer. EBRAINS 2.0 will become the neuroscience hub in the European infrastructure landscape, through building strong links with the European data spaces, EOSC and EuroHPC JU, centers of excellences and other initiatives. Globally, EBRAINS 2.0 will make a strong contribution to the new era of digital neuroscience and foster European leadership in this field.

SafeREG SafeREG project on cordis.europa.eu

Title: Probabilistic Non-Rigid Registration for Safe Brain Tumor Resection

Duration: From August 1, 2024 to July 31, 2026

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France

Inria contact: Demian Wassermann

Coordinator:

Summary: Brain tumors strike people in the prime of life. Surgical resection is the initial treatment for nearly all brain tumors and aims at maximizing the extent of tumor resection while preserving the patient's cognitive function. To optimize this tradeoff, neuronavigation systems have been developed to provide intraoperative guidance to surgeons. These systems allow for the visualization of the position of surgeons' surgical tools relative to the tumor and critical brain areas visible in preoperative Magnetic Resonance Imaging. However, these systems become inaccurate as the surgery progresses since they do not account for brain deformation and tissue resection occurring during surgery.

In an interdisciplinary effort, project SafeREG combines the researcher's background to the expertise of computational scientists from INRIA and clinicians from Parisian hospitals. Its objective is to invent a novel image registration methodology with intraoperative ultrasound that is rich enough to capture complex deformations occurring at the tumor and resection cavity boundaries, fast enough to be employable clinically, and interpretable enough for informed decision-making by neurosurgeons. This will be accomplished by pushing the envelope of scientific knowledge in (1) cross-modality domain adaptation for weakly- and unsupervised image segmentation; (2) modality-invariance representation learning using contrastive learning; (3) non-rigid registration with discrete probabilistic methods; (4) simulated-based variational inference for registration uncertainty quantification that leverages biomechanical knowledge. This research has the potential to deliver accurate and informed image-guided surgery, conferring a lower risk of new neurologic deficits and improved patient prognosis. Beyond neurosurgery, it has broad applications to additional areas of image-guided therapy, including spine, liver, and prostate surgery.

10.3.2 H2020 projects

EBRAIN HEALTH

Title: EBRAIN-HEALTH

Duration: 2023 -> 2026

Coordinator: Petra Ritter (Charité, Berlin)

Partners:

- CHARITE - UNIVERSITAETSMEDIZIN BERLIN
- EBRAINS
- FORSCHUNGSZENTRUM JULICH GMBH
- STICHTING RADBOUD UNIVERSITEIT
- UNIVERSIDAD POMPEU FABRA
- OSLO UNIVERSITETSSYKEHUS HF
- TP21 GMBH
- FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV
- INDOC RESEARCH EUROPE GGMBH
- UNIVERSITAT WIEN
- UNIVERSIDAD COMPLUTENSE DE MADRID
- EODYNE SYSTEMS SL
- ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS
- UNIVERSITETET I OSLO
- STICHTING VUMC
- UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA
- ALZHEIMER EUROPE
- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE

Inria contact: Bertrand Thirion

Summary: The project aims to develop a decentralized, data protection-compliant research platform capable of simulating some of the brain's complex neurobiological phenomena. As part of the project, researchers will collate an array of different types of information, including data from PET and MRI scans, EEG tests, behavioral studies and lifestyle surveys, as well as clinical data from thousands of patients and healthy controls. These will be combined with biological information from knowledge databases and made available for research purposes. The resultant digital "brain twins" will enable large numbers of researchers to conduct innovative research within a powerful digital infrastructure.
Thanks to its transparent analytical pipelines, the new research infrastructure will also help to promote reproducible research.
Furthermore, complex, personalized brain simulations which take into account large quantities of data may be able to provide a better understanding of the mechanisms underlying brain function and disorders. Virtual brain modeling may also improve diagnostics and disease prediction, in addition to enabling the optimization of treatment plans. The project consortium comprises 20 partners and operates in cooperation with EBRAINS AISBL, the coordinating entity of the EU-funded flagship "Human Brain Project".

10.4 National initiatives

CEA Audace at-risk research program

Title: BrainSync

Duration: 11/2024 -> 04/2029

Coordinator: Philippe Ciuciu (CEA/DRF/JOLIOT/NEUROSPIN/MIND), Saclay

Partners:

- CEA/DRF/JOLIOT/NEUROSPIN (BAOBAB, UNICOG), Saclay
- CEA/DRT/LETI/CLINATEC, Grenoble
- CEA/DRT/LIST/DSCIN, Grenoble
- Inria-CEA MIND, Palaiseau
- APHP, FHU NeuroVasc, Paris-Nord
- GHU Paris Neuroscience & Psychiatry
- CHU Grenoble-Alpes
- LPNC, University Grenoble-Alphes & CNRS
- CHU St Etienne

Inria contact: Philippe Ciuciu

Summary: The project aims to understand the learning mechanisms that enable the human brain to adapt to cognitive demand via the flexible recruitment of different regions and connections, thus enabling exploration of an uncertain and changing environment. Understanding these mechanisms in healthy subjects, coupled with the creation of a high-resolution anatomical-functional digital brain atlas of a cohort of post-acute stroke patients (MOTIF-STROKE clinical trial), will enable us to propose AI models predictive of upper limb motor recovery in these patients, and therapeutic innovations (use of neuroprostheses) for a small number of them in a second clinical trial. Clinical evaluation of relearning processes will be based on the use of incremental and adaptive artificial intelligence (AI) algorithms linked to neuroplasticity processes.

This project has received a funding of 5M€ for a period of 4.5 years starting in November 2024 with a go/no-go positioned after a period of 30 months (04/30/2027).

CEA BlueSky project

Title: Brain & Computers

Duration: 2023 -> 2026

Coordinator: Philippe Ciuciu (CEA/DRF/JOLIOT/NEUROSPIN/MIND), Saclay

Partners:

- CEA/DRF/JOLIOT/NEUROSPIN (BAOBAB, UNICOG), Saclay
- CEA/DRT/LETI/CLINATEC, Grenoble
- CEA/DRT/LIST/DSCIN, Grenoble

Inria contact: Philippe Ciuciu

Summary: Artificial Intelligence (AI) is now capable of approaching human performance in tasks such as visual recognition, classification (i.e., decision-making), and even textual or visual production (e.g., GPT-4). Understanding the human brain mechanisms of learning and decision-making in an uncertain environment remains a major scientific challenge in neuroscience. This understanding will enable the development of AI architectures that replicate brain circuits. Additionally, in clinical applications, it can lead to a generational leap in the design of neuroprostheses. These neuroprostheses hold great promise for improving the quality of life for individuals affected by spinal cord injuries (approximately 30% of cases).
The two main objectives of this BlueSky project complement each other. On one hand, in healthy subjects, the goal is to design computational models based on AI that encode these cerebral functions to gain a more precise understanding of the associated brain activity. On the other hand, the objective is to enable a greater number of patients to control neuroprostheses

autonomously through AI, making these medical devices more widely accepted as a therapeutic solution for motor rehabilitation. Tackling such a challenge is not without risks and requires expanding knowledge in neuroscience, surpassing current technological and clinical limits. It also calls for strong synergies among the various CEA institutes involved (Joliot, LETI, and LIST) and their academic (Inserm, Inria, Universities Paris-Saclay, and Grenoble Alpes) and clinical partners (CHU Grenoble-Alpes).

This project has received a funding of 1.5M€ for the 2023-2025 period.

ANR DARLING

Title: DARLING: Distributed adaptation and learning over graph signals

Duration: 2020 -> 2025 (extended)

Coordinator: Cédric Richard (cedric.richard@unice.fr),Professor 3IA Senior Chair in UCA

Partners:

- Université Côte d'Azur Nice, France
- CNRS, École Normale Supérieure, Lyon, France
- Gipsa-lab, UMR 5216, CNRS, UGA, Grenoble, France
- CentraleSupélec, University of Paris-Saclay, Gif-sur-yvette, France

Inria contact: Philippe Ciuciu

Summary: The DARLING project will aim to propose new adaptive learning methods, distributed and collaborative on large dynamic graphs in order to extract structured information of the data flows generated and/or transiting at the nodes of these graphs. In order to obtain performance guarantees, these methods will be systematically accompanied by an in-depth study of random matrix theory. This powerful tool, never exploited so far in this context although perfectly suited for inference on random graphs, will thereby provide even avenues for improvement. Finally, in addition to their evaluation on public data sets, the methods will be compared with each other using two advanced imaging techniques in which two of the partners are involved: radio astronomy with the giant SKA instrument (Obs. Côte d'Azur) and MEG brain imaging (Inria MIND at NeuroSpin, CEA Saclay). Sheng Wang as a postdoc in MIND and Merlin Dumeur as a MIND PhD student in co-tutelle with Matias Palva from Aalto University, Finland are actually involved in the processing of MEG and S/EEG time series on graphs, notably to analyze scale-free (i.e. critical and bistability) phenomena across these graphs and extract potentially new biomarkers for characterizing the pathophysiology of epileptogenic zone (EZ) in drug resistant epilepsy.

ANR VLFMRI

Title: VLFMRI: Very low field MRI for babies

Duration: 2021 -> 2025

Coordinator: Claude Fermon (CEA Saclay, DRF/IRAMIS/SPECT)

Partners:

- CEA/SHFJ/BIOMAPS, Orsay, France
- CEA/NeuroSpin, Gif-sur-Yvette, France
- APHP Robert Debré hospital, Paris, France
- APHP Bicêtre hospital, Kremlin-Bicêtre, France

Inria contact: Philippe Ciuciu

Summary: VLFMRI aims at developing a very low-field Magnetic Resonance Imaging (MRI) system as an alternative to conventional high-field MRI for continuous imaging of premature newborns to detect hemorrhages or ischemia. This system is based on a combination of a new generation of magnetic sensors based on spin electronics, optimized MR acquisition sequences (based on the SPARKLING patent, Inria-CEA MIND team at NeuroSpin) and a open and compatible system with an incubator that will allow to achieve an image resolution of $1mm^3$ on a whole baby body in a short scan time. This project is a partnership of three academic partners and two hospital departments. The different stages of the project are the finalization of the hardware development and software system, preclinical validation on small animals and clinical validation. Kumari Pooja has been hired in January 2022 as research engineer in MIND to interact with the coordinator of this ANR project, Claude Fermon and design new accelerated acquisition methods for verly low field MRI. Preliminary encouraging results allow us to retrospectively accelerate MRI acquisition by a factor of 10 without degrading image quality at 2mm isotropic resolution.

KARAIB AI CHAIR

Title: KARAIB: Knowledge And RepresentAtion Integration on the Brain

Duration: 2020 -> 2024

Coordinator: Bertrand Thirion

Partners:

- INRIA MIND, Gif-sur-Yvette, France

Inria contact: Bertrand Thirion

Summary: Cognitive science describes mental operations, and functional brain imaging provides a unique window into the brain systems that support these operations. A growing body of neuroimaging research has provided significant insight into the relations between psychological functions and brain activity. However, the aggregation of cognitive neuroscience results to obtain a systematic mapping between structure and function faces the roadblock that cognitive concepts are ill-defined and may not map cleanly onto the computational architecture of the brain.

To tackle this challenge, we propose to leverage rapidly increasing data sources: text and brain locations described in neuroscientific publications, brain images and their annotations taken from public data repositories, and several reference datasets. Our aim here is to develop multi-modal machine learning techniques to bridge these data sources.

- Aim 1 develops representation techniques for noisy data to couple brain data with descriptions of behavior or diseases, in order to extract semantic structure.
- Aim 2 challenges these representations to provide explanations to the observed relationships, based on two frameworks: i) a statistical analysis framework; ii) integration into a domain-specific language.
- Aim 3 outputs readily-usable products for neuroimaging: atlases and ontologies and focuses on implementation, with contributions to neuroimaging web-based data sharing tools.site.

BrAIN AI CHAIR

Title: BrAIN: Bridging Artificial Intelligence and Neuroscience

Duration: 2020 -> 2024

Coordinator: Alexandre Gramfort

Partners:

- INRIA MIND, Gif-sur-Yvette, France

Inria contact: Alexandre Gramfort

Summary: The BrAIN project investigates learning tasks from multivariate EEG and MEG time series. In clinical or cognitive neuroscience, electromagnetic signals emitted by synchronously firing neurons are collected by electroencephalography (EEG) or magnetoencephalography (MEG). Such data, typically sampled at millisecond resolution, are routinely used for clinical applications such as anesthesia monitoring, sleep medicine, epilepsy or disorders of consciousness. Low cost EEG devices are also becoming commodities with hardware startups such as DREEM in France or InteraXon in Canada that have collected hundred of thousands of neural recordings. The field of neuroscience urgently needs algorithms that can learn from such large and poorly labeled datasets. The general objectives of BrAIN is to develop ML algorithms that can learn with weak or no supervision on neural time series. It requires contributions to self-supervised learning, domain adaptation and data augmentation techniques, exploiting the known underlying physical mechanisms that govern the data generating process of neurophysiological signals.

The BrAIN project is organized around four objectives:

- Learn with no-supervision on noisy and complex multivariate signals
- Learn end-to-end predictive systems from limited data exploiting physical constraints
- Learn from data coming from many different source domains
- Develop high-quality software tools that can reach clinical research

ANR MICBrainPres

Title: MicBrainPres: Distributed adaptation and learning over graph signals

Duration: 2023 -> 2026

Coordinator: Demian Wassermann

Partners:

- Brain and Spine Institute, Paris, France
- CNRS, Université de Paris, Lyon, France

Inria contact: Demian Wassermann

Summary: The main goal of this project is to harness the latest advances on machine learning-based neuroimage processing technologies to improve function-preserving brain tumour resection. Identifying eloquent brain regions is fundamental to performing tumour resection while preserving a maximum level of cognitive function. Despite the sustained advance in predicting subject-level cognitive abilities from neuroimaging data, current approaches lack sensitivity and specificity in identifying eloquent brain regions. This hinders neuroimaging's usefulness for pre-surgical planning as a tool to predict the preservation of cognitive function after tumour resection. In this project, we propose that using subject-specific parcellations, derived from functional and diffusion MRI through deep-learning technologies, will achieve the needed sensitivity and specificity to locate eloquent areas pre-surgically and to predict cortical reshaping after tumour resection.

ANR EBUL

Title: EBUL: Event-based Unsupervised Learning for Physiological Signals

Duration: 2023 -> 2027

Coordinator: Thomas Moreau

Partners:

- INRIA MIND, Gif-sur-Yvette, France

Inria contact: Thomas Moreau

Summary: Sensor-based body monitoring is now routine clinical care. The resulting records are called physiological signals. While enormous quantities of signals are collected every day, the cost and time necessary to clean and annotate them is prohibitive to constitute large labeled databases. When working with physiological signals, extra sources of information are the events surrounding the recordings. Events are external phenomena that impact the signal and can correlate with the prediction task considered. EBUL propose to develop novel unsupervised learning techniques to process such records based on the notion of events, and to apply them to process general anesthesia records collected in Paris hospital Lariboisière. The methodology of the project relies on the development of novel point process models adapted to capture the distribution of physiological events, and their coupling with event detection algorithms. This will provide novel signal representations based on the distribution of events inside them, which are simpler to analyze and fine tune to derive predictive bio-markers.

The EBUL project is organized around 3 objectives:

- Develop novel point process models for physiological signals
- Learn joint models for signals and events
- Develop high-quality models for general anesthesia that can reach clinical research

ANR BenchArk

Title: BenchArk: An efficient and robust benchmarking suite for AI

Duration: 2024 -> 2028

Coordinator: Thomas Moreau, Mathurin Massias, Joseph Salmon

Partners:

- INRIA MIND, Gif-sur-Yvette, France
- INRIA OCKHAM, Lyon, France
- Université de Montpellier, Montpellier, France

Inria contact: Thomas Moreau

Summary: Numerical evaluation of novel methods, a.k.a. benchmarking, is a pillar of the scientific method in machine learning. However, due to practical and statistical obstacles, the reproducibility of published results is currently insufficient: many details can invalidate numerical comparisons, from insufficient uncertainty quantification to improper methodology. In 2022, the benchopt initiative provided an open source Python package together with a framework to seamlessly run, reuse, share and publish benchmarks in numerical optimization. In this project, we aim at bringing benchopt to the whole machine learning community, making it a new standard in benchmarking by empowering researchers and practitioners with efficient and valid benchmarking methods. Our goal is to ensure reproducibility and consistency in model evaluation. We will federate the machine learning community to develop informative and statistically valid benchmarks, while providing methods to reduce identified hurdles in implementing such practices. The results of the project will be integrated in the open source benchopt library.

Brain Health Trajectories

Title: Brain Health Trajectories

Duration: 2023 -> 2028

Coordinator: Viktor Jirsa, INS Marseille

Partners:

- INSERM Délégation Provence-Alpes-Côte d'Azur et Corse
- CNRS IDF Sud (Gif)
- Université d'Aix-Marseille
- CHU de Grenoble
- CEA, DRF, Joliot, Neurospin
- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE

Inria contact: Bertrand Thirion

Summary: This project is part of PEPR Santé Numérique. The Brain Health Trajectories project is part of the PEPR Santé numérique program, a major research initiative of the French government as part of the France 2030 investment plan.

The project prioritizes the early identification of individuals at risk of developing a disease, so they can receive appropriate treatment before the disease develops or at least adapt their lifestyle to delay the onset of the pathology and slow down the progression of impairment. It aims to develop tools to characterize individuals' brain health and lay the foundation for a platform for screening, decision-making within the population, and prognostic monitoring of treatment efficacy.

11 Dissemination

Participants: Demian Wassermann, Bertrand Thirion, Philippe Ciuciu, Thomas Moreau, Chaithya Giliyar Radhakrishna.

Member of the organizing committees

P. Ciuciu Organizer of the kick-off meeting (40 people on November, 25 2024 at NeuroPSI) of the BrainSync project in the frame of risky research program (Audace!) at CEA. Eight national research insitutions involved (CEA, Inria, UGA, UPSaclay, AP-HP, GHU Paris-Neuroscience, CHU Grenoble-Alpes, CHU St Etienne).

T. Moreau 2019-2023: Co-organizer of the Séminaire Palaisien, a monthly seminar in the Saclay ecosystem around statistics and machine learning (40 people).

T. Moreau Co-organizer of the Learning and optimization in Luminy 2024 workshop, a workshop in the CIRM (Luminy) gathering international researchers working on optimization and learning topics (60 people).

T. Moreau Co-organizer of the Sacl-AI 4 Science workshop, a workshop to gather faculy, researchers and students who work, would like to work or are interested in applications of machine learning to science (100 people). The workshop was composed of 2 introductory days to the machine learning and 3 days of thematic sessions on the applications of machine learning to science.

B.Thirion Organizer of the Workshop on statistical inference for complex data workshop at Inria Saclay.

Reviewer

T. Moreau Reviewer for ICML, NeurIPS 2024 and ICLR 2025.

11.0.1 Journal

Member of the editorial boards

P. Ciuciu Associate Editor (AE) for IEEE Transactions on Medical Imaging (TMI), Senior Area Editor for IEEE Open Journal on Signal Processing, AE for Frontiers in Neuroscience, section Brain Imaging Methods.

B. Thirion Associate Editor (AE) for Medical Image Analysis (MIA) and Transactions on Machine Learning Research.

Reviewer - reviewing activities

P. Ciuciu IEEE Transactions on Medical Imaging, IEEE Transactions on Cloud Computing, Magnetic Resonance in Medicine, NeuroImage.

B. Thirion NeuroImage, MEdIA, IEEE Transactions on Medical Imaging, Brain Structure and Function, Human Brain Mapping, Nature Communications.

C. Giliyar Radhakrishna Magnetic Resonance in Medicine, IEEE Transactions on Medical Imaging.

T. Moreau IEEE Transactions on Signal Processing.

11.0.2 Invited talks

P. Ciuciu

- May 2024: Invited speaker at the Federal University of Sao Paulo (UNIFESP), Sao Paulo, Brazil.
- May 2024: Invited lecture in the educational program at ISMRM 2024, Singapore.
- October 2024: Invited speaker at the School of Medicine, University of Sao Paulo, Sao Paulo, Brazil.

T. Moreau

- April 2024: Invited speaker at the IMCS workshop on Inverse Problems, Edinburgh, UK.
- June 2024: Invited speaker to talk in the ML and brain day, Lyon, France.
- September 2024: Invited to talk in the Noah's Ark internal seminar, Paris, France.
- September 2024: Invited to talk in the Noah's Ark internal seminar, Paris, France.

B.Thirion

- February 2024: Invited speaker at the Montpellier Omics Days (MOD) workshop on Montpellier France.
- March 2024: Plenary talk at AI-MIND European Project General Assembly, Amsterdam, Netherlands.
- October 2024: Plenary talk at Brain and Mind Conference, Helsinki, Finland.
- May 2024: Plenary talk at IMAG Colloquuium, Montpellier, France
- August 2024: Invitation for a seminar at Palva Lab, helsinki, Finland
- December 2024: Plenary talk at PEPR Santé Numérique yearly meeting, Nante, France

11.0.3 Research administration

P. Ciuciu

- Member of the Board of Directors at NeuroSpin (CEA).
- Member of the team leaders committee at Inria Saclay Ile-de-France
- Member of the steering committee of the CEA cross-disciplinary research program on numerical simulation and AI.

B. Thirion

- Délégué Scientifique of Inria Saclay Center
- Member of ENS Paris-Saclay Scientific Council
- Member of Telecom Sud Paris Scientific Council
- Member of Inria Commission évaluation

11.1 Teaching - Supervision - Juries

11.1.1 Teaching

P. Ciuciu

- Tutorial presenter at the 2024 EUSIPCO conference: *Computational MRI in the Deep Learning Era: The Two Facets of Acquisition and Image Reconstruction.*
- Lecturer at the Institut d'Optique Graduate School (3rd year, Signal & Images major).
- Lecturer at the M2 ATSI (CentraleSupelec, ENS Paris-Saclay): Medical imaging course.

D. Wassermann

- Master MVA (École Polytechnique, École Normale Superiore, Centrale Supelec): Graphical Models
- Master in Biomedical Engineering (Université Paris Descartes): Quantification in NeuroImaging.

T. Moreau

- Master Data Science (IPP/UP Saclay): Datacamp.

B.Thirion

- MVA Master (École Polytechnique, École Normale Superiore, Centrale Supelec): Brain Function ; 12h
- NeuroEngineering Master (UPSaclay): fMRI data analysis; 2h

11.1.2 Supervision

P. Ciuciu

- Merlin Dumeur (with M. Palva, Aalto Univ), PhD in cotutelle (4y), 2020-2025 (defense in March 2025)
- Zaineb Amor (with A. Vignaud, CEA) PhD 2020-2024
- Pierre-Antoine Comby (with A. Vignaud, CEA), PhD 2021-2025 (defense in March 2025)
- Serge Brosset, (with Z. Saghi, CEA) PhD 2022-2025
- Dennis Nuñez-Fernandez (with A. Vignaud, CEA), PhD 2023-2024 (PhD stopped in Oct 2024 after resignment)
- Asma Tanabene (with C. Giliyar-Radhakrishna), PhD 2024-2027
- Caini Pan (w. A. Vignaud & C. Giliyar-Radhakrishna), PhD 2024-2027

- Caini Pan, M2, Telecom ParisTech, (Apr - Oct 2024, with C. Giliyar-Radhakrishna)
- Benjamin Lapostolle, M2 MVA BME Paris, (Apr - Aug 2024, with M. Terris)

B. Thirion

- Alexis Thual, PhD 2020-2024 (with S. Dehaene, INSERM), defended on June 13th, 2024
- Ahmad Chamma, PhD 2021-2024 (with D.Engemann, Roche), defended on June 14th, 2024
- Raphael Meudec, PhD 2021-2024 (with D.Wassermann, Inria), defended on June 18th, 2024
- Thomas Chapalain, PhD 2021-2025 (with E.Eger, CEA)
- Alexandre Blain, PhD 2021-2024 (with P.Neuvial, CNRS), defended on December 9th, 2024
- Nicolas Salvy, PhD 2023-2026 (with H.Talbot, CentraleSupelec)
- Joseph Paillard, 2024-2027 (with D.Engemann, Roche)
- Angel Reyero Lobo, 2024-2027 (with P.Neuvial, CNRS)
- Sonia Mazelet, 2024-2027 (with R.Flamary, IPParis)
- Pierre-Louis Barabarant, 2024-2027 (with F.Meyniel, CEA)
- Fernanda Ponce, 2024-2027 (with D.Wassermann, Inria)

D. Wassermann

- Gaston Zanitti, PhD 2020-2023
- Louis Rouillard, PhD 2021-2024
- Raphael Meudec, PhD 2021-2024 (with B. Thirion)
- Alexandre Le Bris, PhD 2022-2025
- Gabriela Gomez Jimenez, PhD 2023-2026 (with J. Valette CEA)

T. Moreau

- C. Allain (with A. Gramfort, Meta), PhD 2021-2024
- M. Dagréou (with S. Vaiter, Université Cote d'Azur and P. Ablin, Apple), PhD 2021-2024
- F. Michel (with M. Kowalski, UP Saclay), PhD 2022-2024 (stopped due to personal reasons)
- J. Perdereau (with F. Vallée, APHP), PhD 2022-2025
- V. Loison (with J. Cartailler, APHP), PhD 2023-2026
- C. Eve (with G. Varoquaux, Inria), PhD 2024-2027

11.1.3 Juries

B. Thirion

- Examiner for Robin Louiset's PhD defense on June 19th, Saclay, France
- Reveiwer for Axel Kroner(s PhD defense on Sept 9th, Maastricht, Netherlands
- Eaminer for Pietro Gori's habilitation defense on Sept 5th, Paris, France
- Examiner for Shizhe Wu's private defense on Nov 6th, Geneva, Switzerland
- Examiner for Marua Sayu Yamamoto's PhD defense on Dec. 2nd, Palaiseau, France
- Opponent for Vladislav Myrov' PhD defense on Dec. 13th , Helsinki, Finland
- Examiner for Sacha Haudry's PhD defense on Dec 20th, Caen, France

11.2 Popularization

11.2.1 Specific official responsibilities in science outreach structures

D. Wassermann COERLE Scientific representative for Inria Saclay Île-de-France; Representative at the Graduate School in Computer Science of Université Paris-Saclay for Inria Saclay Île-de-France.

T. Moreau President of the Inria Saclay CUMI; Representative for Inria Saclay in the commission for the development of the national computational resources, member of the user's comittee for Jean-zay high performance computing.

11.2.2 Productions (articles, videos, podcasts, serious games, ...)

P.Ciuciu Participation in the NeuroJET event at NeuroSpin on Saturday November, 30 to advertise internship projects to Master students.

P. Ciuciu CEA video produced during the Digital Mission Days in November 2024 to present the BlueSky project. Broadcasted at the CEA DRT General Assembly on Jan 30, 2025.

P. Ciuciu CEA video produced for the kick-off meeting of the BrainSync project on November, 25 2024 and then disseminated on LinkedIn. Broadcasted at the CEA DRF General Assembly on Jan 23, 2025.

12 Scientific production

12.1 Major publications

[1] M. Abdallah, G. Zanitti, V. Iovene and D. Wassermann. 'Functional gradients in the human lateral prefrontal cortex revealed by a comprehensive coordinate-based meta-analysis'. In: *eLife* (2022). DOI: 10.7554/eLife.76926. URL: https://hal.inria.fr/hal-03544221.

[2] Z. Amor, C. Le Ster, C. Gr, G. Daval-Frérot, N. Boulant, F. Mauconduit, B. Thirion, P. Ciuciu and A. Vignaud. 'Impact of B0 field imperfections correction on BOLD sensitivity in 3D-SPARKLING fMRI data'. In: *Magnetic Resonance in Medicine* (29th Dec. 2023). DOI: 10.1002/mrm.29943. URL: https://hal.science/hal-04363381.

[3] C. Caucheteux, A. Gramfort and J.-R. King. 'Deep language algorithms predict semantic comprehension from brain activity'. In: *Scientific Reports* (29th Sept. 2022). DOI: 10.1038/s41598-022-20460-9. URL: https://hal.archives-ouvertes.fr/hal-03361426.

[4] M. Dagréou, P. Ablin, S. Vaiter and T. Moreau. 'A framework for bilevel optimization that enables stochastic and global variance reduction algorithms'. In: Advances in Neural Information Processing Systems (NeurIPS). New Orleans, United States, 28th Nov. 2022. URL: https://hal.archives-ouvertes.fr/hal-03562151.

[5] C. Giliyar Radhakrishna, G. Daval-Frérot, A. Massire, A. Vignaud and P. Ciuciu. 'Improving spreading projection algorithm for rapid k-space sampling trajectories through minimized off-resonance effects and gridding of low frequencies'. In: *Magnetic Resonance in Medicine* 90.3 (22nd May 2023), pp. 1069–1085. DOI: 10.1002/mrm.29702. URL: https://inria.hal.science/hal-04370475 (cit. on p. 25).

[6] C. Giliyar Radhakrishna, P. Weiss, G. Daval-Frérot, A. Massire, A. Vignaud and P. Ciuciu. 'Optimizing full 3D SPARKLING trajectories for high-resolution T2*-weighted Magnetic Resonance Imaging'. In: *IEEE Transactions on Medical Imaging* (8th Aug. 2022). URL: https://hal.inria.fr/hal-03090471 (cit. on p. 25).

[7] T. Moreau, M. Massias, A. Gramfort, P. Ablin, P.-A. Bannier, B. Charlier, M. Dagréou, T. Dupré La Tour, G. Durif, C. F. Dantas, Q. Klopfenstein, J. Larsson, E. Lai, T. Lefort, B. Malézieux, B. Moufad, B. T. Nguyen, A. Rakotomamonjy, Z. Ramzi, J. Salmon and S. Vaiter. 'Benchopt: Reproducible, efficient and collaborative optimization benchmarks'. In: NeurIPS 2022 - 36th Conference on Neural Information Processing Systems. New Orleans, United States, 28th Nov. 2022. URL: https://hal.archives-ouvertes.fr/hal-03830604.

[8] Z. Ramzi, C. Giliyar Radhakrishna, J.-L. Starck and P. Ciuciu. 'NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction'. In: *IEEE Transactions on Medical Imaging* (20th Jan. 2022). DOI: 10.1109/TMI.2022.3144619. URL: https://hal.inria.fr/hal-03188997.

[9] C. Rommel, J. Paillard, T. Moreau and A. Gramfort. 'Data augmentation for learning predictive models on EEG: a systematic comparison'. In: *Journal of Neural Engineering* (11th Nov. 2022). DOI: 10.1088/1741-2552/aca220. URL: https://hal.archives-ouvertes.fr/hal-03853329.

[10] G. Staerman, C. Allain, A. Gramfort and T. Moreau. 'FaDIn: Fast Discretized Inference for Hawkes Processes with General Parametric Kernels'. In: *PMLR*. International Conference on Machine Learning. Vol. 202. Honolulu, Hawaii, United States, 25th July 2023, pp. 32575–32597. URL: https://hal.science/hal-03811810.

12.2 Publications of the year

International journals

[11] Z. Amor, P. Ciuciu, C. G. R., G. Daval-Frérot, F. Mauconduit, B. Thirion and A. Vignaud. 'Non-cartesian 3D-SPARKLING vs cartesian 3D-EPI encoding schemes for functional magnetic resonance Imaging at 7 Tesla'. In: *PLoS ONE* 19.5 (13th May 2024), e0299925. DOI: 10.1371/journal.pone.0299925. URL: https://hal.science/hal-04350075 (cit. on p. 26).

[12] Z. Amor, C. Le Ster, C. Giliyar Radhakrishna, G. Daval-Frérot, N. Boulant, F. Mauconduit, B. Thirion, P. Ciuciu and A. Vignaud. 'Impact of B0 field imperfections correction on BOLD sensitivity in 3D-SPARKLING fMRI data'. In: *Magnetic Resonance in Medicine* Volume 91.Issue 4 (1st Apr. 2024), Pages 1434–1448. DOI: 10.1002/mrm.29943. URL: https://hal.science/hal-04363381 (cit. on p. 26).

[13] M. Barberis, I. Poisson, C. Prévost-Tarabon, S. Letrange, S. Froelich, B. Thirion and E. Mandonnet. 'Verbal fluency predicts work resumption after awake surgery in low-grade glioma patients'. In: *Acta Neurochirurgica* 166.1 (19th Feb. 2024), p. 88. DOI: 10.1007/s00701-024-05971-w. URL: https://hal.science/hal-04712003.

[14] V. Chirokoff, S. Berthoz, M. Fatseas, D. Misdrahi, M. Dupuy, M. Abdallah, F. Serre, M. Auriacombe, A. Pfefferbaum, E. V. Sullivan and S. Chanraud. 'Identifying the role of (dis)inhibition in the vicious cycle of substance use through ecological momentary assessment and resting-state fMRI'. In: *Translational Psychiatry* 14.1 (2024), p. 260. DOI: 10.1038/s41398-024-02949-1. URL: https://hal.science/hal-04804003.

[15] G. Hamon, L. Legrand, G. Hmeydia, G. Turc, W. B. Hassen, S. Charron, C. Debacker, O. Naggara, B. Thirion, B. Chen, B. Lapergue, C. Oppenheim and J. Benzakoun. 'Multicenter Validation of Synthetic FLAIR as a substitute for FLAIR Sequence in Acute Ischemic Stroke'. In: *European Stroke Journal* (3rd Aug. 2024), p. 23969873241263418. DOI: 10.1177/23969873241263418. URL: https://inserm.hal.science/inserm-04667549.

[16] G. J. M. van der Lande, D. Casas-Torremocha, A. Manasanch, L. Dalla Porta, O. Gosseries, N. Alnagger, A. Barra, J. F. Mejías, R. Panda, F. Riefolo, A. Thibaut, V. Bonhomme, B. Thirion, F. Clasca, P. Gorostiza, M. V. Sanchez-Vives, G. Deco, S. Laureys, G. Zamora-López and J. Annen. 'Brain state identification and neuromodulation to promote recovery of consciousness'. In: *Brain Communications* 6.5 (2nd Sept. 2024). DOI: 10.1093/braincomms/fcae362. URL: https://inria.hal.science/hal-04851113.

[17] V. Pacella, V. Nozais, L. Talozzi, M. Abdallah, D. Wassermann, S. Forkel and M. Thiebaut de Schotten. 'The morphospace of the brain-cognition organisation'. In: *Nature Communications* 15.1 (30th Sept. 2024), p. 8452. DOI: 10.1038/s41467-024-52186-9. URL: https://hal.science/hal-04714643.

[18] A. Pérez-Millan, B. Thirion, N. Falgàs, S. Borrego-Écija, B. Bosch, J. Juncà-Parella, A. Tort-Merino, J. Sarto, J. M. Augé, A. Antonell, N. Bargalló, M. Balasa, A. Lladó, R. Sánchez-Valle and R. Sala-Llonch. 'Beyond group classification: Probabilistic differential diagnosis of frontotemporal dementia and Alzheimer's disease with MRI and CSF biomarkers'. In: *Neurobiology of Aging* 144 (Dec. 2024), pp. 1–11. DOI: 10.1016/j.neurobiolaging.2024.08.008. URL: https://inria.hal.science/hal-04851068.

[19] A. L. Pinho, H. Richard, A. F. Ponce, M. Eickenberg, A. Amadon, E. Dohmatob, I. Denghien, J. J. Torre, S. Shankar, H. Aggarwal, A. Thual, T. Chapalain, C. Ginisty, S. Becuwe-Desmidt, S. Roger, Y. Lecomte, V. Berland, L. Laurier, V. Joly-Testault, G. Médiouni-Cloarec, C. Doublé, B. Martins, G. Varoquaux, S. Dehaene, L. Hertz-Pannier and B. Thirion. 'Individual Brain Charting dataset extension, third release for movie watching and retinotopy data'. In: *Scientific Data* 11.590 (5th June 2024). URL: https://hal.science/hal-04272993.

[20] R. Poldrack, C. Markiewicz, S. Appelhoff, Y. Ashar, T. Auer, S. Baillet, S. Bansal, L. Beltrachini, C. Benar, G. Bertazzoli et al. 'The Past, Present, and Future of the Brain Imaging Data Structure (BIDS)'. In: *Imaging Neuroscience* 2 (2024), pp. 1–19. DOI: 10.1162/imag_a_00103. URL: https://hal.science/hal-04346097.

[21] Z. Saghi, L. Guetaz, T. David and P. Ciuciu. 'Deep image prior for limited-angle electron tomography'. In: *BIO Web of Conferences* 129 (17th Oct. 2024), p. 02012. DOI: 10.1051/bioconf/202412902012. URL: https://hal.science/hal-04747438.

[22] V. Shevchenko, R. A. Benn, R. Scholz, W. Wei, C. Pallavicini, U. Klatzmann, F. Alberti, T. D. Satterthwaite, D. Wassermann, P.-L. Bazin and D. S. Margulies. 'A comparative machine learning study of schizophrenia biomarkers derived from functional connectivity'. In: *Scientific Reports* 15.2849 (22nd Jan. 2025). DOI: 10.1038/s41598-024-84152-2. URL: https://hal.science/hal-04911960.

[23] N. Tolley, P. L. Coelho Rodrigues, A. Gramfort and S. R. Jones. 'Methods and considerations for estimating parameters in biophysically detailed neural models with simulation based inference'. In: *PLoS Computational Biology* 20.2 (26th Feb. 2024), e1011108. DOI: 10.1371/journal.pcbi.1011108. URL: https://hal.science/hal-04856130.

[24] C. Truong and T. Moreau. 'Convolutional Sparse Coding for Time Series Via a l0 Penalty: An Efficient Algorithm With Statistical Guarantees'. In: *Statistical Analysis and Data Mining* 17.6 (16th Dec. 2024), e70000. DOI: 10.1002/sam.70000. URL: https://hal.science/hal-04856979.

[25] H.-T. Wang, S. Meisler, H. Sharmarke, N. Clarke, N. Gensollen, C. Markiewicz, F. Paugam, B. Thirion and P. Bellec. 'Continuous evaluation of denoising strategies in resting-state fMRI connectivity using fMRIPrep and Nilearn'. In: *PLoS Computational Biology* 20.3 (18th Mar. 2024), e1011942. DOI: 10.1371/journal.pcbi.1011942. URL: https://hal.science/hal-04718572.

[26] S. H. Wang, G. Arnulfo, L. Nobili, V. Myrov, P. Ferrari, P. Ciuciu, S. Palva and J. M. Palva. 'Neuronal synchrony and critical bistability: Mechanistic biomarkers for localizing the epileptogenic network'. In: *Epilepsia* 65.7 (30th Apr. 2024), pp. 2041–2053. DOI: 10.1111/epi.17996. URL: https://hal.science/hal-04843581.

International peer-reviewed conferences

[27] H. Aggarwal, L. Al-Shikhley and B. Thirion. 'Across-subject ensemble-learning alleviates the need for large samples for fMRI decoding'. In: *Lecture Notes in Computer Science (LNCS, volume 15003)*. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, 27th International Conference, Proceedings, Part III. Vol. LNCS 15003. Lecture Notes

in Computer Science (LNCS, volume 15003). Marakkech, Morocco, Nov. 2024, pp. 1–11. URL: https://hal.science/hal-04636523.

[28] M. Dagréou, T. Moreau, S. Vaiter and P. Ablin. 'A lower bound and a near-optimal algorithm for bilevel empirical risk minimization'. In: International Conference on Artificial Intelligence and Statistics (AISTATS). Valencia, Spain, 2024. DOI: 10.48550/arXiv.2302.08766. URL: https://hal.science/hal-04302861.

[29] M. Darrin, G. Staerman, E. Dadalto Camara Gomes, J. Cheung, P. Piantanida and P. Colombo. 'Unsupervised Layer-Wise Score Aggregation for Textual OOD Detection'. In: *Proceedings of the 38th Conference on Artificial Intelligence*. AAAI 2024 - 38th Conference on Artificial Intelligence. Vol. 38. 16. Vancouver, Canada, 24th Mar. 2024, pp. 17880–17888. DOI: 10.1609/aaai.v38i16.29742. URL: https://hal.science/hal-04575132.

[30] D. Degras, T. Chapalain and B. Thirion. 'On the Stability of Reduced-Rank Ridge Regression'. In: EUSIPCO 2024 - 32nd European signal processing conference. Lyon, France, 26th Aug. 2024. URL: https://inria.hal.science/hal-04720738.

[31] G. Gómez Jiménez and D. Wassermann. 'Deep multivariate autoencoder for capturing complexity in Brain Structure and Behaviour Relationships'. In: CDMRI 2024 - International Workshop on Computational Diffusion MRI. Marrakech, Morocco, 6th Oct. 2024. URL: https://inria.hal.science/hal-04680736.

[32] A. Mellot, A. Collas, S. Chevallier, D. Engemann and A. Gramfort. 'Physics-informed and Unsupervised Riemannian Domain Adaptation for Machine Learning on Heterogeneous EEG Datasets'. In: EUSIPCO 2024 - 32nd European Signal Processing Conference. Lyon, France: arXiv, 2024. DOI: 10.48550/arXiv.2403.15415. URL: https://hal.science/hal-04716030.

[33] A. Mellot, A. Collas, S. Chevallier, A. Gramfort and D. Engemann. 'Geodesic optimization for predictive shift adaptation on EEG data'. In: *Proc. NeuIPS'24*. NeurIPS 2024 - 38th Conference on Neural Information Processing Systems. Vancouver, Canada, 10th Dec. 2024. URL: https://hal.science/hal-04902523.

[34] R. Meudec, F. Ghayem, J. Dockes, D. Wassermann and B. Thirion. 'NeuroConText: Contrastive text-to-brain mapping for neuroscientific literature'. In: MICCAI24 - 27th International conference on medical image computing and computer assisted intervention. Marrakesh, Morocco, 6th Oct. 2024. URL: https://hal.science/hal-04708173 (cit. on p. 28).

Conferences without proceedings

[35] Z. Amor, P.-A. Comby, P. Ciuciu and A. Vignaud. 'Achieving high temporal resolution using a sliding-window approach for SPARKLING fMRI data: A simulation study'. In: ISMRM & ISMRT 2024 - Annual Meeting & Exhibition. Singapore, Singapore, 4th May 2024. URL: https://hal.science/hal-04427082.

[36] S. H. Wang, M. Marzulli, G. Arnulfo, L. Nobili, S. Palva, J. M. Palva and P. Ciuciu. 'Machine Learning Models Trained in a Low-Dimensional Latent Space for Epileptogenic Zone (EZ) Localization'. In: 2024 32nd European Signal Processing Conference (EUSIPCO). Lyon, France: IEEE, 26th Aug. 2024, pp. 1586–1590. DOI: 10.23919/EUSIPCO63174.2024.10715112. URL: https://hal.science/hal-04843593.

Scientific book chapters

[37] B. Thirion. 'Machine learning for NeuroImaging data analysis'. In: *Encyclopedia of the Human Brain*. Elsevier, 2025, pp. 580–588. DOI: 10.1016/B978-0-12-820480-1.00158-3. URL: https://inria.hal.science/hal-04901990.

Doctoral dissertations and habilitation theses

[38] C. Allain. 'Temporal point processes and scalable convolutional dictionary learning : a unified framework for m/eeg signal analysis in neuroscience'. Université Paris-Saclay, 9th Feb. 2024. URL: https://theses.hal.science/tel-04611761.

[39] A. Chamma. 'Statistical interpretation of high-dimensional complex prediction models for biomedical data'. Université Paris-Saclay, 14th June 2024. URL: https://theses.hal.science/tel-04619339.

[40] M. Dagréou. 'Contributions to stochastic bilevel optimization'. Université Paris-Saclay, 8th Oct. 2024. URL: https://theses.hal.science/tel-04877851.

[41] J. Linhart. 'Simulation-based inference with deep learning : application to neuroscience time series data'. Université Paris-Saclay, 5th Dec. 2024. URL: https://theses.hal.science/tel-04920410.

[42] A. Mellot. 'Machine learning and domain adaptation for enhancing the measure of brain health with MEG and EEG signals'. Université Paris-Saclay, 8th Nov. 2024. URL: https://theses.hal.science/tel-04906458.

[43] L. Rouillard. 'Bridging Simulation-based Inference and Hierarchical Modeling : Applications in Neuroscience'. Université Paris-Saclay, 3rd May 2024. URL: https://theses.hal.science/tel-04612209.

Reports & preprints

[44] P. Abry, P. Ciuciu, M. Dumeur, S. Jaffard and G. Saës. *Multifractal analysis based on the weak scaling exponent and applications to MEG recordings in neuroscience*. 3rd Nov. 2024. URL: https://hal.science/hal-04712582.

[45] M. Bertrait, C. Giliyar Radhakrishna and P. Ciuciu. *gGRAPPA: A Flexible, GPU-Accelerated Python Package for Fast and Efficient generalized GRAPPA Reconstruction*. 22nd Jan. 2025. URL: https://inria.hal.science/hal-04906124 (cit. on p. 27).

[46] A. Blain, B. Thirion, J. Linhart and P. Neuvial. *When Knockoffs fail: diagnosing and fixing non-exchangeability of Knockoffs*. 9th July 2024. URL: https://hal.science/hal-04642238.

[47] S. Chevallier, I. Carrara, B. Aristimunha, P. Guetschel, S. Sedlar, B. Junqueira Lopes, S. Velut, S. Khazem and T. Moreau. *The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark*. 8th Apr. 2024. URL: https://universite-paris-saclay.hal.science/hal-04537061.

[48] A. Collas, R. Flamary and A. Gramfort. *Weakly supervised covariance matrices alignment through Stiefel matrices estimation for MEG applications*. 28th Jan. 2024. URL: https://hal.science/hal-04422094.

[49] P.-A. Comby, G. Daval-Frérot, C. Pan, A. Tanabene, L. Oudjman, M. Cencini, P. Ciuciu and C. Giliyar Radhkrishna. *MRI-NUFFT: Doing non-Cartesian MRI has never been easier*. 9th Nov. 2024. URL: https://hal.science/hal-04775262.

[50] P.-A. Comby, B. Lapostolle, M. Terris and P. Ciuciu. *Robust plug-and-play methods for highly accelerated non-Cartesian MRI reconstruction*. 24th Jan. 2025. URL: https://hal.science/hal-04759015.

[51] P.-A. Comby, A. Vignaud and P. Ciuciu. *SNAKE-fMRI: A modular fMRI data simulator from the space-time domain to k-space and back*. 29th Mar. 2024. URL: https://hal.science/hal-04533862 (cit. on p. 27).

[52] C. Giliyar Radhakrishna, A. Vignaud, M. Bertrait, A. Massire, M. Bottlaender and P. Ciuciu. *Bringing GRAPPA to non-Cartesian MRI through SPARKLING: An application to MPRAGE anatomical MRI*. 14th Jan. 2025. URL: https://hal.science/hal-04885977 (cit. on p. 25).

[53] H. Haque, S. Wang, F. Siebenhühner, E. Robertson, J. M. Palva and S. Palva. *Synchronization networks reflect the contents of visual working memory*. 12th Jan. 2024. DOI: 10.21203/rs.3.rs-3853906/v1. URL: https://hal.science/hal-04843624.

[54] A. Himmi, G. Staerman, M. Picot, P. Colombo and N. Guerreiro. *Enhanced Hallucination Detection in Neural Machine Translation through Simple Detector Aggregation*. 14th May 2024. URL: https://hal.science/hal-04574891.

[55] M. Kowalski, B. Malézieux, T. Moreau and A. Repetti. *Analysis and synthesis denoisers for forward-backward plug-and-play algorithms*. 11th Dec. 2024. URL: https://hal.science/hal-04786802.

[56] A. Le Bris, L. Rouillard and D. Wassermann. *Improving Individual-Specific Functional Parcellation Through Transfer Learning*. 8th Mar. 2024. URL: https://hal.science/hal-04506561.

[57] J. Linhart, G. V. Cardoso, A. Gramfort, S. Le Corff and P. L. Coelho Rodrigues. *Diffusion posterior sampling for simulation-based inference in tall data settings*. 7th June 2024. URL: https://hal.science/hal-04459545.

[58] B. Malézieux, F. Michel, T. Moreau and M. Kowalski. *Where prior learning can and can't work in unsupervised inverse problems*. 14th Nov. 2024. URL: https://hal.science/hal-04782335.

[59] A. Tanabene, C. Giliyar Radhakrishna, A. Massire, M. S. Nadar and P. Ciuciu. *Benchmarking 3D multi-coil NC-PDNET MRI reconstruction*. 31st Oct. 2024. URL: https://inria.hal.science/hal-04771143.

[60] M. Terris, U. S. Kamilov and T. Moreau. *FiRe: Fixed-points of restoration priors for solving inverse problems*. 28th Nov. 2024. URL: https://hal.science/hal-04814326.

[61] J. Zhou, P. Gaillard, T. Rahier, H. Zenati and J. Arbel. *Towards Efficient and Optimal Covariance-Adaptive Algorithms for Combinatorial Semi-Bandits*. 2024. URL: https://hal.science/hal-04470568.

Other scientific publications

[62] B. Aristimunha, T. Moreau, S. Chevallier, R. Y. de Camargo and M.-C. Corsi. 'What is the best model for decoding neurophysiological signals? Depends on how you evaluate'. In: CNS 2024 - 33rd Annual Computational Neuroscience Meeting. Natal, Brazil, 20th July 2024. URL: https://inria.hal.science/hal-04743845.

[63] M. Dagréou, P. Ablin, S. Vaiter and T. Moreau. *How to compute Hessian-vector products?* 7th May 2024. URL: https://hal.science/hal-04869111.

[64] F. Paul, W. Sheng H, T. Moreau, A. Gabriele, N. Lino, P. Satu, P. Ciuciu and P. Matias. 'Non-invasive MEG Localization of Excitatory and Inhibitory Spikes Using Convolutional Dictionary Learning'. In: American Society for Epilepsy Meeting (AES 2024). Los angeles, United States, 6th Dec. 2024. URL: https://hal.science/hal-04843738.

[65] W. Sheng H., M. Morgane, D. Degras, A. Gabriele, N. Lino, M. Vladislav, P. Satu, F. Paul, P. Matias and P. Ciuciu. 'A Low-dimensional Assessments of Seizure Risk Dynamics'. In: American Society for Epilepsy Meeting. Los Ageles, United States, 6th Dec. 2024. URL: https://hal.science/hal-04843718.

12.3 Cited publications

[66] M. Abadi. In: (cit. on p. 7).

[67] P. Ablin, J.-F. Cardoso and A. Gramfort. 'Faster independent component analysis by preconditioning with Hessian approximations'. In: *IEEE Trans. Signal Process.* 66.15 (2018), pp. 4040–4049 (cit. on p. 13).

[68] P. Ablin, T. Moreau, M. Massias and A. Gramfort. 'Learning Step Sizes for Unfolded Sparse Coding'. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vancouver, BC, Canada, 2019, pp. 13100–13110. arXiv: 1905.11071 (cit. on p. 7).

[69] P. Ablin, G. Peyré and T. Moreau. 'Super-Efficiency of Automatic Differentiation for Functions Defined as a Minimum'. en. In: *International Conference on Machine Learning (ICML)*. July 2020 (cit. on p. 7).

[70] A. Alaa and M. Van Der Schaar. 'Validating Causal Inference Models via Influence Functions'. In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. Long Beach, California, USA: PMLR, Sept. 2019, pp. 191–201. URL: http://proceedings.mlr.press/v97/alaa19a.html (cit. on p. 12).

[71] M. Arenas, G. Gottlob and a. Pieris. 'Expressive Languages for Querying the Semantic Web'. en. In: *ACM Transactions on Database Systems* 43.3 (Nov. 2018), pp. 1–45. DOI: 10.1145/3238304 (cit. on p. 9).

[72] S. Athey and G. Imbens. 'The State of Applied Econometrics - Causality and Policy Evaluation'. In: *ArXiv e-prints*, arXiv:1607.00699 (July 2016), arXiv:1607.00699. arXiv: 1607.00699 [stat.ME] (cit. on p. 11).

[73] S. Aydore, B. Thirion and G. Varoquaux. 'Feature Grouping as a Stochastic Regularizer for High-Dimensional Structured Data'. In: *ICML*. 2019, pp. 385–394 (cit. on p. 9).

[74] T. Baltrušaitis, C. Ahuja and L. Morency. 'Multimodal Machine Learning: A Survey and Taxonomy'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2 (Feb. 2019), pp. 423–443. DOI: 10.1109/TPAMI.2018.2798607 (cit. on pp. 9, 16).

[75] H. Banville, O. Chehab, A. Hyvarinen, D. Engemann and A. Gramfort. 'Uncovering the structure of clinical EEG signals with self-supervised learning'. In: *Journal of Neural Engineering* (2020) (cit. on pp. 13, 18).

[76] V. BáRány, B. T. Cate, B. Kimelfeld, D. Olteanu and Z. Vagena. 'Declarative Probabilistic Programming with Datalog'. en. In: *ACM Transactions on Database Systems* 42.4 (Oct. 2017), pp. 1–35. DOI: 10.1145/3132700 (cit. on p. 10).

[77] L. Bellomarini, E. Sallinger and G. Gottlob. 'The Vadalog System: Datalog-Based Reasoning for Knowledge Graphs'. en. In: *Proceedings of the VLDB Endowment* 11.9 (May 2018), pp. 975–987. DOI: 10.14778/3213880.3213888 (cit. on p. 10).

[78] Y. Bengio, A. Courville and P. Vincent. 'Representation learning: A review and new perspectives'. In: *Pattern Analysis and Machine Intelligence* 35 (2013), p. 1798 (cit. on p. 16).

[79] Y. Benjamini and Y. Hochberg. 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing'. In: *Journal of the Royal Statistical Society Series B (Methodological)* 57.1 (1995), pp. 289–300. DOI: \url{http://dx.doi.org/10.2307/2346101}. URL: http://dx.doi.org/10.2307/2346101 (cit. on p. 11).

[80] Q. Bertrand, Q. Klopfenstein, M. Blondel, S. Vaiter, A. Gramfort and J. Salmon. 'Implicit Differentiation of Lasso-Type Models for Hyperparameter Optimization'. en. In: *International Conference on Machine Learning (ICML)*. Vol. 2002.08943. online, Apr. 2020, pp. 3199–3210. arXiv: 2002.08943 (cit. on p. 7).

[81] G. Blanchard, P. Neuvial and E. Roquain. 'Post hoc confidence bounds on false positives using reference families'. In: *Ann. Statist.* 48.3 (June 2020), pp. 1281–1303. DOI: 10.1214/19-AOS1847. URL: https://doi.org/10.1214/19-AOS1847 (cit. on p. 11).

[82] M. Bompaire. 'Machine learning based on Hawkes processes and stochastic optimization'. fr. PhD thesis. CMAP, École Polytechnique: Université Paris Saclay (COmUE), 2019 (cit. on p. 14).

[83] C. Boyer, N. Chauffert, P. Ciuciu, J. Kahn and P. Weiss. 'On the generation of sampling schemes for magnetic resonance imaging'. In: *SIAM Journal on Imaging Sciences* 9.4 (2016), pp. 2039–2072 (cit. on p. 8).

[84] E. Candès, Y. Fan, L. Janson and J. Lv. 'Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection'. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.3 (2018), pp. 551–577. DOI: \url{https://doi.org/10.1111/rssb.12265}. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12265 (cit. on p. 10).

[85] L. Chaari, P. Ciuciu, S. Mériaux and J.-C. Pesquet. 'Spatio-temporal wavelet regularization for parallel MRI reconstruction: application to functional MRI'. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 27.6 (2014), pp. 509–529 (cit. on p. 6).

[86] G. Chaithya, P. Weiss, A. Massire, A. Vignaud and P. Ciuciu. 'Globally optimized 3D SPARKLING trajectories for high-resolution T2*-weighted Magnetic Resonance Imaging'. In: (2020) (cit. on pp. 8, 19).

[87] S. Chambon, V. Thorey, P. J. Arnal, E. Mignot and A. Gramfort. 'A Deep Learning Architecture to Detect Events in EEG Signals During Sleep'. In: *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. Sept. 2018, pp. 1–6. DOI: 10.1109/MLSP.2018.8517067 (cit. on p. 13).

[88] N. Chauffert, P. Weiss, J. Kahn and P. Ciuciu. 'A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging'. In: *IEEE Transactions on Medical Imaging* 35.9 (2016), pp. 2026–2039 (cit. on p. 8).

[89] L. Chen, D. Wassermann, D. A. Abrams, J. Kochalka, G. Gallardo-Diez and V. Menon. 'The Visual Word Form Area (VWFA) Is Part of Both Language and Attention Circuitry'. en. In: *Nature Communications* 10.1 (Dec. 2019), p. 5601. DOI: 10.1038/s41467-019-13634-z (cit. on p. 14).

[90] H. Cherkaoui, T. Moreau, A. Halimi, C. Leroy and P. Ciuciu. 'Multivariate semi-blind deconvolution of fMRI time series'. In: *revised for publication to NeuroImage* (Apr. 2021) (cit. on p. 13).

[91] H. Cherkaoui, L. El Gueddari, C. Lazarus, A. Grigis, F. Poupon, A. Vignaud, S. Farrens, J.-L. Starck and P. Ciuciu. 'Analysis vs synthesis-based regularization for combined compressed sensing and parallel MRI reconstruction at 7 Tesla'. In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 36–40 (cit. on p. 6).

[92] H. Cherkaoui, J. Sulam and T. Moreau. 'Learning to Solve TV Regularised Problems with Unrolled Algorithms'. en. In: *Advances in Neural Information Processing Systems (NeurIPS)*. online, 2020 (cit. on p. 7).

[93] J.-A. Chevalier, J. Salmon and B. Thirion. 'Statistical Inference with Ensemble of Clustered Desparsified Lasso'. In: *MICCAI*. Grenade, Spain, 2018. URL: https://hal.archives-ouvertes.fr/hal-01815255 (cit. on p. 12).

[94] J. H. Cole, R. Leech, D. J. Sharp and A. D. N. Initiative. 'Prediction of brain age suggests accelerated atrophy after traumatic brain injury'. In: *Annals of neurology* 77.4 (2015), pp. 571–581 (cit. on p. 15).

[95] R. Collobert and J. Weston. 'A unified architecture for natural language processing: Deep neural networks with multitask learning'. In: *ICML*. 2008, p. 160 (cit. on p. 16).

[96] E. D. Cubuk, B. Zoph, J. Shlens and Q. V. Le. 'Randaugment: Practical automated data augmentation with a reduced search space'. en. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, WA, USA: IEEE, June 2020, pp. 3008–3017. DOI: 10.1109/CVPRW50498.2020.00359. URL: https://ieeexplore.ieee.org/document/9150790/ (visited on 15/12/2020) (cit. on p. 13).

[97] K. Dadi, G. Varoquaux, J. Houenou, D. Bzdok, B. Thirion and D. Engemann. 'Beyond brain age: Empirically-derived proxy measures of mental health'. working paper or preprint. Oct. 2020. DOI: 10.1101/2020.08.25.266536. URL: https://hal.inria.fr/hal-02929857 (cit. on p. 16).

[98] K. Dadi, G. Varoquaux, A. Machlouzarides-Shalit, K. J. Gorgolewski, D. Wassermann, B. Thirion and A. Mensch. 'Fine-grain atlases of functional modes for fMRI analysis'. In: *NeuroImage* 221 (2020), p. 117126 (cit. on p. 28).

[99] J. M. Danskin. *Theory of Max-Min and Its Application to Weapons Allocation Problems*. en. Berlin/Heidelberg: Springer Berlin Heidelberg, 1967 (cit. on p. 7).

[100] G. Daval-Frérot, A. Massire, M. Ripart, B. Mailhe, M. Nadar, A. Vignaud and P. Ciuciu. 'Off-resonance correction non-Cartesian SWI using internal field map estimation'. In: *29th Proc. of the ISMRM annual meeting*. virtual, May 2021 (cit. on p. 6).

[101] J. Deng, A. C. Berg, K. Li and L. Fei-Fei. 'What does classifying more than 10,000 image categories tell us?' In: *ECCV*. 2010, p. 71 (cit. on p. 16).

[102] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova. 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. In: *CoRR* abs/1810.04805 (2018). arXiv: 1810.04805. URL: http://arxiv.org/abs/1810.04805 (cit. on p. 16).

[103] E. Dohmatob, H. Richard, A. L. Pinho and B. Thirion. 'Brain topography beyond parcellations: local gradients of functional maps'. In: *NeuroImage* (Jan. 2021), p. 117706. DOI: 10.1016/j.neuroimage.2020.117706. URL: https://hal.inria.fr/hal-03122173 (cit. on p. 14).

[104] N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar et al. 'Prediction of individual brain maturity using fMRI'. In: *Science* 329.5997 (2010), pp. 1358–1361 (cit. on p. 15).

[105] T. Dupré la Tour, T. Moreau, M. Jas and A. Gramfort. 'Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals'. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Montreal, Canada, 2018, pp. 3296–3306 (cit. on p. 13).

[106] L. El Gueddari, E. Chouzenoux, A. Vignaud, J.-C. Pesquet and P. Ciuciu. 'Online MR image reconstruction for compressed sensing acquisition in T2* imaging'. In: *Wavelets and Sparsity XVIII*. Vol. 11138. International Society for Optics and Photonics. 2019, p. 1113819 (cit. on p. 6).

[107] L. El Gueddari, C. Giliyar Radhakrishna, E. Chouzenoux and P. Ciuciu. 'Calibration-Less Multi-Coil Compressed Sensing Magnetic Resonance Image Reconstruction Based on OSCAR Regularization'. In: *Journal of Imaging* 7.3 (2021), p. 58 (cit. on p. 6).

[108] L. El Gueddari, C. Lazarus, H. Carrié, A. Vignaud and P. Ciuciu. 'Self-calibrating nonlinear reconstruction algorithms for variable density sampling and parallel reception MRI'. In: *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE. 2018, pp. 415–419 (cit. on p. 6).

[109] D. A. Engemann, O. Kozynets, D. Sabbagh, G. Lemaître, G. Varoquaux, F. Liem and A. Gramfort. 'Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers'. In: *eLife* 9 (May 2020). Ed. by A. Shackman, F. P. de Lange, K. Tsvetanov and N. Trujillo-Barreto, e54055. DOI: 10.7554/eLife.54055. URL: https://doi.org/10.7554/eLife.54055 (cit. on pp. 15, 18).

[110] D. A. Engemann, F. Raimondo, J.-R. King, B. Rohaut, G. Louppe, F. Faugeras, J. Annen, H. Cassol, O. Gosseries, D. Fernandez-Slezak et al. 'Robust EEG-based cross-site and cross-protocol classification of states of consciousness'. In: *Brain* 141.11 (2018), pp. 3179–3192 (cit. on p. 17).

[111] S. Farrens, A. Grigis, L. El Gueddari, Z. Ramzi, G. Chaithya, S. Starck, B. Sarthou, H. Cherkaoui, P. Ciuciu and J.-L. Starck. 'PySAP: Python Sparse Data Analysis Package for Multidisciplinary Image Processing'. In: *Astronomy and Computing* 32 (2020), p. 100402 (cit. on pp. 6, 19).

[112] D. Gilton, G. Ongie and R. Willett. 'Deep Equilibrium Architectures for Inverse Problems in Imaging'. In: *arXiv preprint arXiv:2102.07944* (2021) (cit. on p. 7).

[113] J. Gonneaud, A. T. Baria, A. P. Binette, B. A. Gordon, J. P. Chhatwal, C. Cruchaga, M. Jucker, J. Levin, S. Salloway, M. Farlow et al. 'Functional brain age prediction suggests accelerated aging in preclinical familial Alzheimer's disease, irrespective of fibrillar amyloid-beta pathology'. In: *bioRxiv* (2020) (cit. on p. 15).

[114] K. J. Gorgolewski, G. Varoquaux, G. Rivera, Y. Schwarz, S. S. Ghosh, C. Maumet, V. V. Sochat, T. E. Nichols, R. A. Poldrack, J.-B. Poline et al. 'NeuroVault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain'. In: *Frontiers in neuroinformatics* 9 (2015), p. 8 (cit. on p. 28).

[115] K. Gregor and Y. Lecun. 'Learning Fast Approximations of Sparse Coding'. In: *Proceedings of the 27th International Conference on Machine Learning*. 2010 (cit. on p. 7).

[116] A. G. Hawkes. 'Point spectra of some mutually exciting point processes'. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 33.3 (1971), pp. 438–443 (cit. on p. 14).

[117] A. A. Hoyos-Idrobo, G. Varoquaux, Y. Schwartz and B. Thirion. 'FReM – scalable and stable decoding with fast regularized ensemble of models'. In: *NeuroImage* (2017), pp. 1–16. DOI: 10.1016/j.neuroimage.2017.10.005. URL: https://hal.archives-ouvertes.fr/hal-01615015 (cit. on p. 9).

[118] V. Iovene, G. Zanitti and D. Wassermann. 'Complex Coordinate-Based Meta-Analysis with Probabilistic Programming'. In: *Association for the Advancement of Artificial Intelligence*. Online, France, Feb. 2021. URL: https://hal.inria.fr/hal-03036125 (cit. on p. 9).

[119] M. Jallais, P. L. Coelho Rodrigues, A. Gramfort and D. Wassermann. 'Cytoarchitecture Measurements in Brain Gray Matter using Likelihood-Free Inference'. In: Rø, June 2021. URL: https://hal.inria.fr/hal-03090959 (cit. on p. 6).

[120] H. Janati, T. Bazeille, B. Thirion, M. Cuturi and A. Gramfort. 'Group Level MEG/EEG Source Imaging via Optimal Transport: Minimum Wasserstein Estimates'. In: *Information Processing in Medical Imaging*. Ed. by A. C. S. Chung, J. C. Gee, P. A. Yushkevich and S. Bao. Cham: Springer International Publishing, 2019, pp. 743–754 (cit. on p. 6).

[121] H. Janati, M. Cuturi and A. Gramfort. 'Wasserstein regularization for sparse multi-task regression'. In: *AISTATS*. Ed. by K. Chaudhuri and M. Sugiyama. Vol. 89. Proceedings of Machine Learning Research. PMLR, Apr. 2019, pp. 1407–1416. URL: http://proceedings.mlr.press/v89/janati19a.html (cit. on p. 6).

[122] M. Jas, T. Dupré la Tour, U. Şimşekli and A. Gramfort. 'Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding'. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Long Beach, CA, USA, 2017, pp. 1099–1108. arXiv: 1705.08006 (cit. on p. 13).

[123] A. Javanmard and A. Montanari. 'Confidence Intervals and Hypothesis Testing for High-Dimensional Regression'. In: 15.1 (Jan. 2014), pp. 2869–2909 (cit. on p. 10).

[124] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier et al. 'Mistral 7B'. In: *arXiv preprint arXiv:2310.06825* (2023) (cit. on p. 28).

[125] J. Josse, N. Prost, E. Scornet and G. Varoquaux. 'On the consistency of supervised learning with missing values'. working paper or preprint. Mar. 2019. URL: https://hal.archives-ouvertes.fr/hal-02024202 (cit. on p. 15).

[126] S. Kapur, A. G. Phillips and T. R. Insel. 'Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it?' In: *Molecular Psychiatry* 17.12 (2012), pp. 1174–1179 (cit. on p. 15).

[127] J.-R. King, L. Gwilliams, C. Holdgraf, J. Sassenhagen, A. Barachant, D. Engemann, E. Larson and A. Gramfort. 'Encoding and Decoding Framework to Uncover the Algorithms of Cognition'. In: *The Cognitive Neurosciences VI, MIT Press* in press (2020) (cit. on p. 17).

[128] E. Lashgari, D. Liang and U. Maoz. 'Data augmentation for deep-learning-based electroencephalography'. In: *Journal of Neuroscience Methods* 346 (2020), p. 108885. DOI: \url{https://doi.org/10.1016/j.jneumeth.2020.108885}. URL: https://www.sciencedirect.com/science/article/pii/S0165027020303083 (cit. on p. 12).

[129] C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, L. El Gueddari, C. Destrieux, I. Zemmoura, A. Vignaud and P. Ciuciu. 'SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI'. In: *Magnetic Resonance in Medicine* 81.6 (2019), pp. 3643–3661 (cit. on pp. 8, 25).

[130] C. Lazarus, P. Weiss, L. Gueddari, F. Mauconduit, A. Massire, M. Ripart, A. Vignaud and P. Ciuciu. '3D variable-density SPARKLING trajectories for high-resolution T2*-weighted Magnetic Resonance imaging'. In: *NMR in Biomedicine* 33.e4349 (2020), pp. 1–12 (cit. on pp. 8, 19, 25).

[131] H. K. Lee, J.-H. Lee, J.-O. Park and Y.-S. Choi. 'Data-driven Data Augmentation for Motor Imagery Brain-Computer Interface'. In: *2021 International Conference on Information Networking (ICOIN)*. ISSN: 1976-7684. Jan. 2021, pp. 683–686. DOI: 10.1109/ICOIN50884.2021.9333908 (cit. on p. 12).

[132] F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. K. Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock et al. 'Predicting brain-age from multimodal imaging data captures cognitive impairment'. In: *NeuroImage* 148 (2017), pp. 179–188 (cit. on p. 15).

[133] M. Liu, E. Katsevich, L. Janson and A. Ramdas. *Fast and Powerful Conditional Randomization Testing via Distillation*. 2020. arXiv: 2006.03980 [stat.ME] (cit. on pp. 10, 13).

[134] F. Lotte. 'Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain-Computer Interfaces'. In: *Proc. of the IEEE* 103.6 (2015), pp. 871–890 (cit. on p. 12).

[135] C. Louizos, U. Shalit, J. Mooij, D. Sontag, R. Zemel and M. Welling. 'Causal Effect Inference with Deep Latent-Variable Models'. In: *arXiv e-prints*, arXiv:1705.08821 (May 2017), arXiv:1705.08821. arXiv: 1705.08821 [stat.ML] (cit. on p. 12).

[136] J. Mairal, F. R. Bach, J. Ponce and G. Sapiro. 'Online Learning for Matrix Factorization and Sparse Coding'. In: *Journal of Machine Learning Research (JMLR)* 11.1 (2010), pp. 19–60. arXiv: 0908.0050 (cit. on p. 7).

[137] M. Massias, A. Gramfort and J. Salmon. 'Celer: a Fast Solver for the Lasso with Dual Extrapolation'. In: *Proceedings of the 35th International Conference on Machine Learning*. Vol. 80. 2018, pp. 3321–3330. URL: https://arxiv.org/abs/1802.07481 (cit. on p. 6).

[138] M. Massias, S. Vaiter, A. Gramfort and J. Salmon. 'Dual Extrapolation for Sparse GLMs'. In: *Journal of Machine Learning Research* 21.234 (2020), pp. 1–33. URL: http://jmlr.org/papers/v21/19-587.html (cit. on p. 6).

[139] V. Menon, G. Gallardo, M. A. Pinsk, V.-D. Nguyen, J.-R. Li, W. Cai and D. Wassermann. 'Microstructural Organization of Human Insula Is Linked to Its Macrofunctional Circuitry and Predicts Cognitive Control'. en. In: *eLife* 9 (June 2020), e53470. DOI: 10.7554/eLife.53470 (cit. on p. 14).

[140] A. Mensch, J. Mairal, D. Bzdok, B. Thirion and G. Varoquaux. 'Learning Neural Representations of Human Cognition across Many fMRI Studies'. In: *NIPS*. Long Beach, United States, Dec. 2017. URL: https://hal.archives-ouvertes.fr/hal-01626823 (cit. on p. 9).

[141] A. Mensch, J. Mairal, B. Thirion and G. Varoquaux. 'Stochastic Subsampling for Factorizing Huge Matrices'. en. In: *IEEE Transactions on Signal Processing* 66.1 (Jan. 2018), pp. 113–128 (cit. on p. 14).

[142] I. Misra, C. L. Zitnick and M. Hebert. 'Shuffle and learn: unsupervised learning using temporal order verification'. In: *ECCV*. Springer. 2016, pp. 527–544 (cit. on p. 13).

[143] T. Moreau and A. Gramfort. 'DiCoDiLe: Distributed Convolutional Dictionary Learning'. In: *Transaction on Pattern Analysis and Machine Intelligence (TPAMI)* in press (2020) (cit. on pp. 13, 14).

[144] T. Moreau, L. Oudre and N. Vayatis. 'DICOD: Distributed Convolutional Sparse Coding'. In: *International Conference on Machine Learning (ICML)*. Stockohlm, Sweden: PMLR (80), 2018, pp. 3626–3634 (cit. on pp. 13, 14).

[145] S. A. Neymotin, D. S. Daniels, B. Caldwell, R. A. McDougal, N. T. Carnevale, M. Jas, C. I. Moore, M. L. Hines, M. Hämäläinen and S. R. Jones. 'Human Neocortical Neurosolver (HNN), a new software tool for interpreting the cellular and network origin of human MEG/EEG data'. In: *eLife* 9 (Jan. 2020). Ed. by R. B. Ivry, A. Stolk, A. Stolk and S. S. Dalal, e51214. DOI: 10.7554/eLife.51214. URL: https://doi.org/10.7554/eLife.51214 (cit. on p. 20).

[146] T.-B. Nguyen, J.-A. Chevalier and B. Thirion. 'ECKO: Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data'. In: *IPMI*. June 2019. URL: https://hal.archives-ouvertes.fr/hal-02076510 (cit. on p. 12).

[147] T.-B. Nguyen, J.-A. Chevalier, B. Thirion and S. Arlot. 'Aggregation of Multiple Knockoffs'. In: *ICML 2020 - 37th International Conference on Machine Learning*. Proceedings of the ICML 37th International Conference on Machine Learning, 119. Vienne / Virtual, Austria, July 2020. URL: https://hal.archives-ouvertes.fr/hal-02888693 (cit. on p. 10).

[148] X. Nie and S. Wager. 'Quasi-Oracle Estimation of Heterogeneous Treatment Effects'. In: *ArXiv e-prints*, arXiv:1712.04912 (Dec. 2017), arXiv:1712.04912. arXiv: 1712.04912 [stat.ML] (cit. on p. 11).

[149] F. Niu, C. Zhang, C. Ré and J. W. Shavlik. 'DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference.' In: *VLDS*. 2012, p. 25 (cit. on p. 16).

[150] M. Noroozi and P. Favaro. 'Unsupervised learning of visual representations by solving jigsaw puzzles'. In: *ECCV*. Springer. 2016, pp. 69–84 (cit. on p. 13).

[151] A. v. d. Oord, Y. Li and O. Vinyals. 'Representation learning with contrastive predictive coding'. In: *arXiv e-prints* (2018) (cit. on p. 13).

[152] L. Paninski. *Lecture notes in Statistical analysis of neural data course: Chapter 2 – Introduction to Point Processes*. Sept. 2019 (cit. on p. 14).

[153] M. Paschali, W. Simson, A. Guha Roy, M. Ferjad Naeem, R. Göbl, C. Wachinger and N. Navab. 'Data Augmentation with Manifold Exploring Geometric Transformations for Increased Performance and Robustness'. In: *arXiv e-prints* (Jan. 2019). arXiv: 1901.04420 (cit. on p. 12).

[154] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala. 'PyTorch: An Imperative Style, High-Performance Deep Learning Library'. en. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vancouver, BC, Canada, 2019, p. 12 (cit. on p. 7).

[155] L. Pineda et al. 'Active MR k-space Sampling with Reinforcement Learning'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12262 LNCS (2020), pp. 23–33 (cit. on p. 8).

[156] A. L. Pinho, A. Amadon, M. Fabre, E. Dohmatob, I. Denghien, J. J. Torres, C. Ginisty, S. Becuwe-Desmidt, S. Roger, L. Laurier, V. Joly-Testault, G. Médiouni-Cloarec, C. Doublé, B. Martins, P. Pinel, E. Eger, G. Varoquaux, C. C. Pallier, S. Dehaene, L. Hertz-Pannier and B. Thirion. 'Subject-specific segregation of functional territories based on deep phenotyping'. In: *Human Brain Mapping* (Dec. 2020). DOI: 10.1002/hbm.25189. URL: https://hal.archives-ouvertes.fr/hal-03088606 (cit. on p. 14).

[157] A. L. Pinho, A. Amadon, B. Gauthier, N. Clairis, A. Knops, S. Genon, E. Dohmatob, J. Jesús Torre, C. Ginisty, S. Becuwe-Desmidt, S. Roger, Y. Lecomte, V. Berland, L. Laurier, V. Joly-Testault, G. Médiouni-Cloarec, C. Doublé, B. Martins, E. Salmon, M. Piazza, D. Melcher, M. Pessiglione, V. Van Wassenhove, E. Eger, G. Varoquaux, S. Dehaene, L. Hertz-Pannier and B. Thirion. 'Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping'. In: *Scientific Data* 7.1 (Oct. 2020). DOI: 10.1038/s41597-020-00670-4. URL: https://hal.archives-ouvertes.fr/hal-02973046 (cit. on p. 14).

[158] Z. Ramzi, J.-L. Starck and P. Ciuciu. 'XPDNet for MRI Reconstruction: an application to the 2020 fastMRI challenge'. In: *ISMRM*. 2021 (cit. on pp. 6, 19).

[159] Z. Ramzi, P. Ciuciu and J.-L. Starck. 'Density Compensated Unrolled Networks for Non-Cartesian MRI Reconstruction'. In: *ISBI 2021 - International Symposium on Biomedical Imaging*. 2021. arXiv: 2101.01570. URL: http://arxiv.org/abs/2101.01570 (cit. on pp. 6, 19).

[160] Z. Ramzi, C. Giliyar Radhakrishna, J.-L. Starck and P. Ciuciu. 'NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction'. In: *IEEE Transactions on Medical Imaging* (Jan. 2022). This work is an extended version of the work presented at the 2021 ISBI conference. DOI: 10.1109/TMI.2022.3144619. URL: https://inria.hal.science/hal-03188997 (cit. on p. 26).

[161] P. L. C. Rodrigues, T. Moreau, G. Louppe and A. Gramfort. 'HNPE: Leveraging Global Parameters for Neural Posterior Estimation'. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Dec. 2021 (cit. on p. 6).

[162] J. D. Rosenblatt, L. Finos, W. D. Weeda, A. Solari and J. J. Goeman. 'All-Resolutions Inference for brain imaging'. In: *Neuroimage* 181 (Nov. 2018), pp. 786–796 (cit. on p. 11).

[163] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk and J. Faubert. 'Deep learning-based electroencephalography analysis: a systematic review'. In: *Journal of Neural Engineering* 16.5 (Aug. 2019), p. 051001. DOI: 10.1088/1741-2552/ab260c. URL: https://doi.org/10.1088/1741-2552/ab260c (cit. on p. 13).

[164] D. B. Rubin. 'Estimating causal effects of treatments in randomized and nonrandomized studies.' In: *Journal of educational Psychology* 66.5 (1974), p. 688 (cit. on pp. 11, 12).

[165] D. Sabbagh, P. Ablin, G. Varoquaux, A. Gramfort and D. A. Engemann. 'Manifold-regression to predict from MEG/EEG brain signals without source modeling'. In: *arXiv preprint arXiv:1906.02687* (2019) (cit. on p. 18).

[166] D. Sabbagh, P. Ablin, G. Varoquaux, A. Gramfort and D. A. Engemann. 'Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states'. In: *NeuroImage* 222 (2020), p. 116893 (cit. on pp. 15, 18).

[167] P. Sanz Leon, S. Knock, M. Woodman, L. Domide, J. Mersmann, A. McIntosh and V. Jirsa. 'The Virtual Brain: a simulator of primate brain network dynamics'. In: *Frontiers in Neuroinformatics* 7 (2013), p. 10. DOI: 10.3389/fninf.2013.00010. URL: https://www.frontiersin.org/article/10.3389/fninf.2013.00010 (cit. on p. 20).

[168] D. W. Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015 (cit. on p. 28).

[169] B. W. Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018 (cit. on p. 28).

[170] S. M. Smith, D. Vidaurre, F. Alfaro-Almagro, T. E. Nichols and K. L. Miller. 'Estimation of brain age delta from brain imaging'. In: *NeuroImage* (2019) (cit. on p. 15).

[171] D. Spiegelhalter. 'How old are you, really? Communicating chronic risk through 'effective age'of your body and organs'. In: *BMC medical informatics and decision making* 16.1 (2016), pp. 1–6 (cit. on p. 15).

[172] D. Suciu. 'Probabilistic Databases for All'. en. In: *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. Portland OR USA: ACM, June 2020, pp. 19–31. DOI: 10.1145/3375395.3389129 (cit. on p. 9).

[173] J. Vennekens, M. Denecker and M. Bruynooghe. 'CP-Logic: A Language of Causal Probabilistic Events and Its Relation to Logic Programming'. en. In: *Theory and Practice of Logic Programming* 9.3 (May 2009), pp. 245–308. DOI: 10.1017/S1471068409003767 (cit. on p. 10).

[174] D. Wassermann, N. Makris, Y. Rathi, M. Shenton, R. Kikinis, M. Kubicki and C.-F. Westin. 'The White Matter Query Language: A Novel Approach for Describing Human White Matter Anatomy.' In: *Brain Structure and Function* (Jan. 2016). DOI: 10.1007/s00429-015-1179-4 (cit. on p. 10).

[175] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang and M. Drozdzal. 'Reducing Uncertainty in Undersampled MRI Reconstruction with Active Acquisition'. In: (Feb. 2019). arXiv: 1902.03051. URL: http://arxiv.org/abs/1902.03051 (cit. on p. 8).

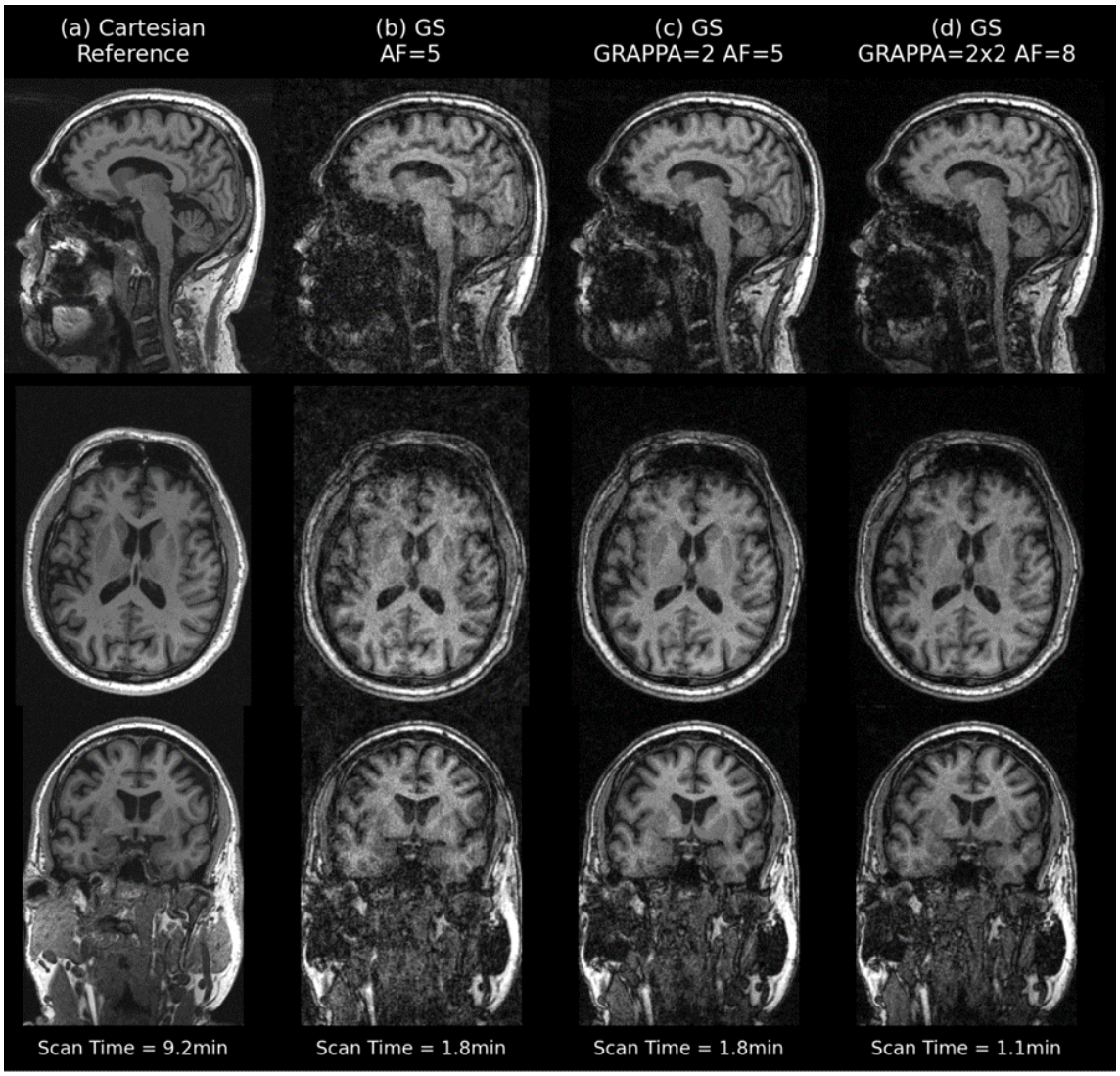


Figure 1: **Accelerated T1-weighted anatomical MRI using GRAPPA-GoLF-SPARKLING**. Reconstructed images for the SENIOR cohort collected at NeuroSpin (3T) with MPRAGE sequence at 1mm isotropic resolution for different GS without GRAPPA (b), with GRAPPA 2x1 acceleration (c) and GRAPPA 2x2 acceleration (d). The corresponding acceleration factor (AF) and the scan times are noted at the top and bottom, respectively. The fully sampled Cartesian reference is shown in (a).

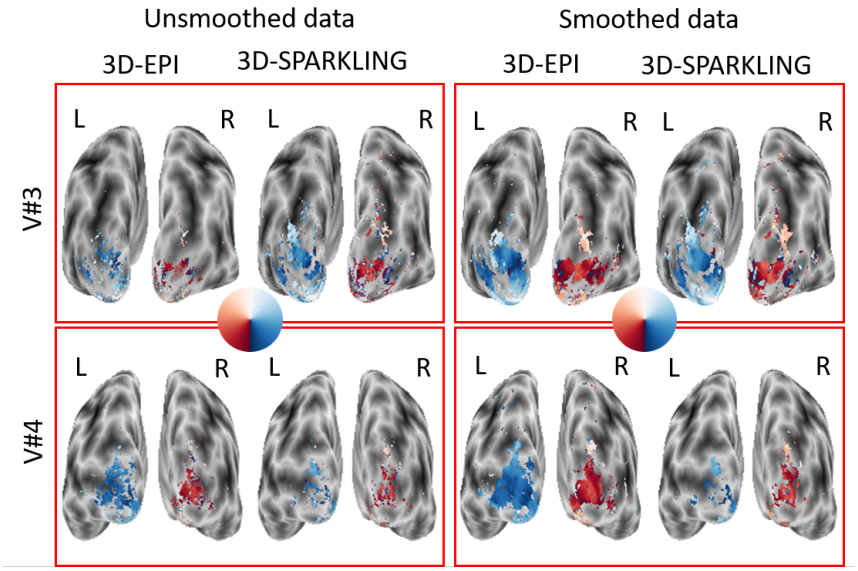


Figure 2: **Projection of the BOLD phase maps on the pial surface visualized on the inflated surface for participants V#3 (3D-SPARKLING run first) and V#4 (3D-EPI run first)**. 3D-SPARKLING yields improved projected BOLD phase maps for V#3 in comparison with 3D-EPI both on raw and spatially smoothed data. Opposite results were found in favor of 3D-EPI in V#4, notably on spatially smoothed data.

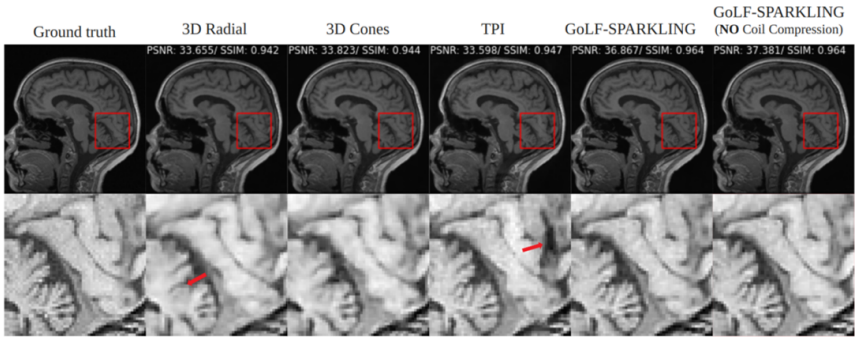


Figure 3: **3D multicoil NCPDNet MRI reconstruction.** Reconstruction results of the 90th slice of file e14079s3 P09216.7 from the test set in Calgary-Campinas dataset. The top row shows reconstructions using different methods, while the bottom row displays zoomed-in regions outlined by red frames. Volume-wise PSNR and SSIM scores are indicated at the top of each image.

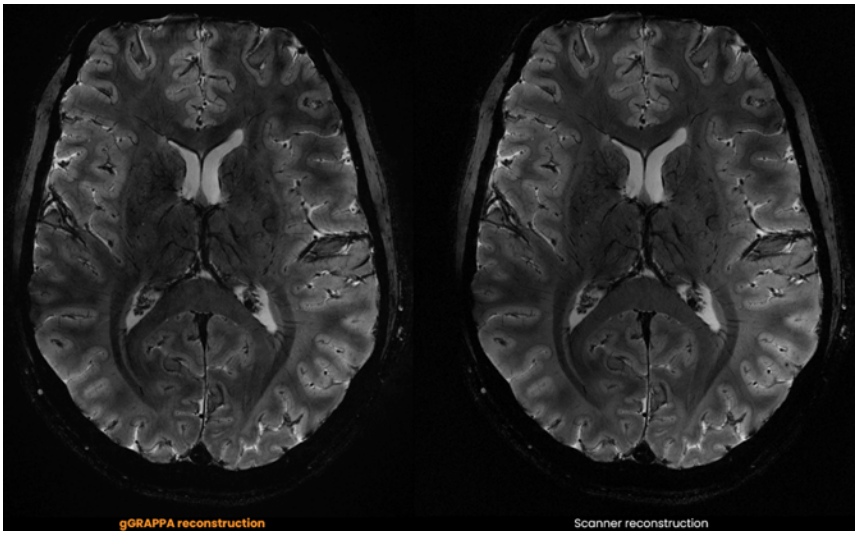


Figure 4: **Ultra-high resolution in vivo brain MRI at 11.7T.** Slice of a T2*-weighted 2D GRE scan at 11.7T from a healthy volunteer (approved by the national ethical committee and ANSM, the French medical device regulatory authority), with a resolution of 1024×1024 at 0.2 mm isotropic and GRAPPA 2×1. The scan is reconstructed by gGRAPPA (left) and the scanner (right), both using identical GRAPPA parameters: a kernel size of 3×4 and a regularization strength of 1e-4. Comparable reconstruction quality is observed between gGRAPPA and the scanner.

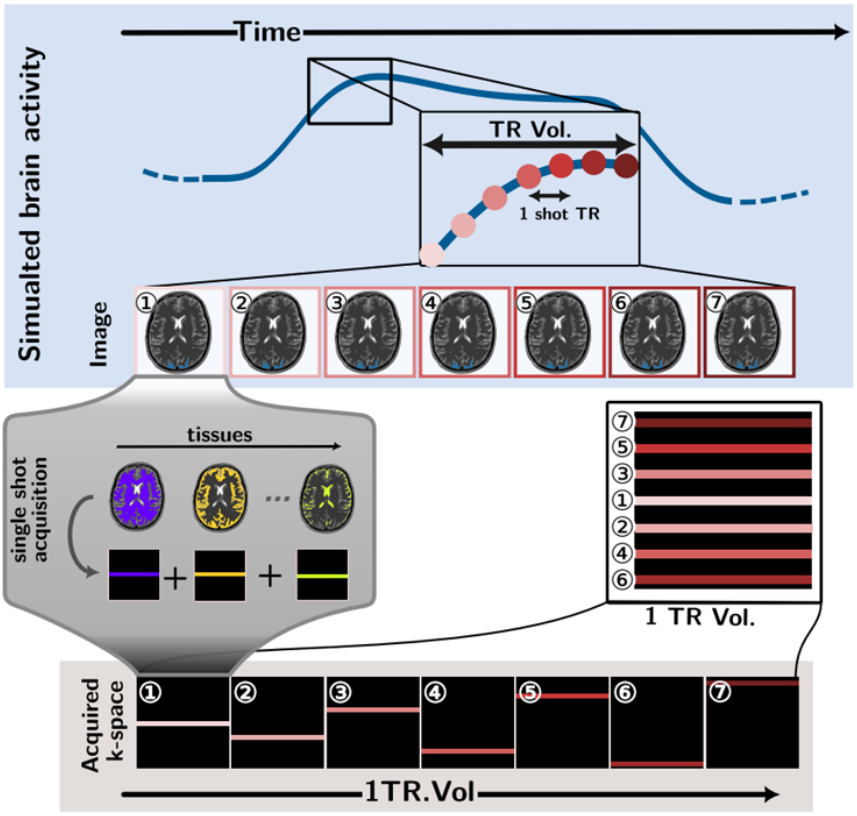


Figure 5: **SNAKE-fMRI simulator.** Acquisition method implemented in SNAKE – The case represented is simplified to a 2D Cartesian case (e.g., a projected view of a 3D non-accelerated EPI scheme). Each shot (i.e., a plane in 3D EPI) of the k-space sampling pattern is acquired separately from an on-the-fly simulated volume in the image domain, as shown in the blue frame. The shots are numbered here from 1 to 7. The acquisition is performed in parallel for each tissue type to apply the T2* relaxation model.

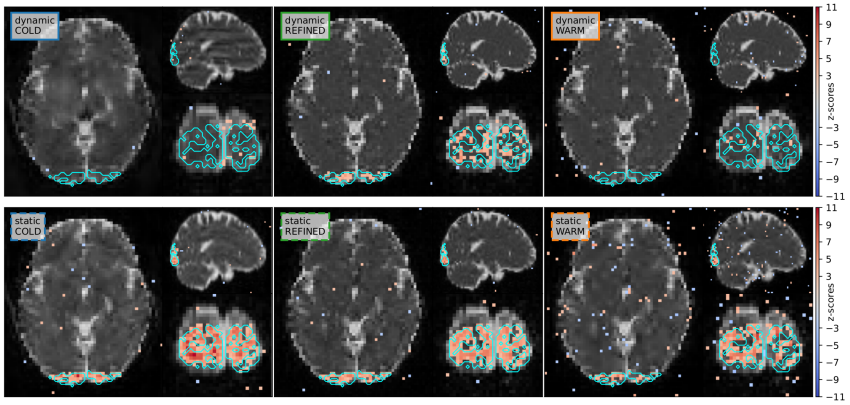
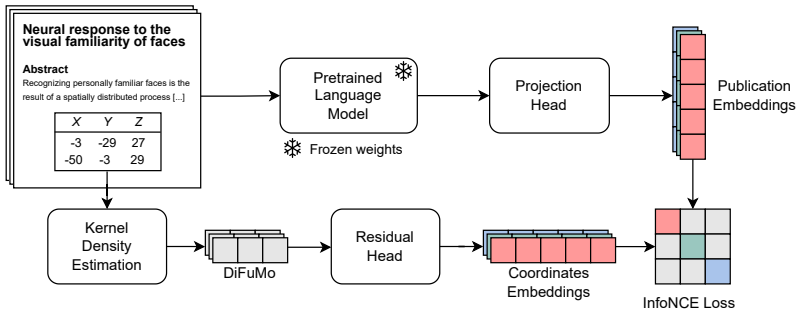
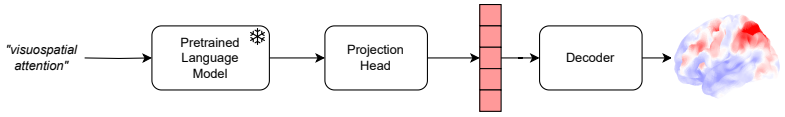


Figure 6: **Activations maps superimposed on the mean fMRI image in the high-temporal (0.7s) imaging setup using stack-of-spiral sampling and CS reconstruction.** Top row: activation maps where a tiume-varying sampling pattern has been used. Bottom row: static sampling pattern, scan and repeat strategy. From left to right: different image reconstruction strategies (COLD, REFINED, WARM), which correspond to different initialization in a framewise CS reconstruction method. Detected activations surviving at p 0.001, uncorrected, thresholding are overlaid using a colorjet map.



(a) Training of NeuroConText contrastive model on neuroscientific publications.



(b) Brain encoding of a query through the NeuroConText text latent representation.

Figure 7: **NeuroConText:** (a) We train a contrastive model on a large corpus to retrieve a shared latent space between coordinates and text from neuroscientific articles. Pre-trained LLMs are used to obtain an initial text embedding, and a projection layer aligns this embedding with those of coordinates. Snowflakes denote models with frozen weights. (b) A decoder is trained from the text latent space to reproduce brain images from any query, enabling the mapping of queries into brain representations.

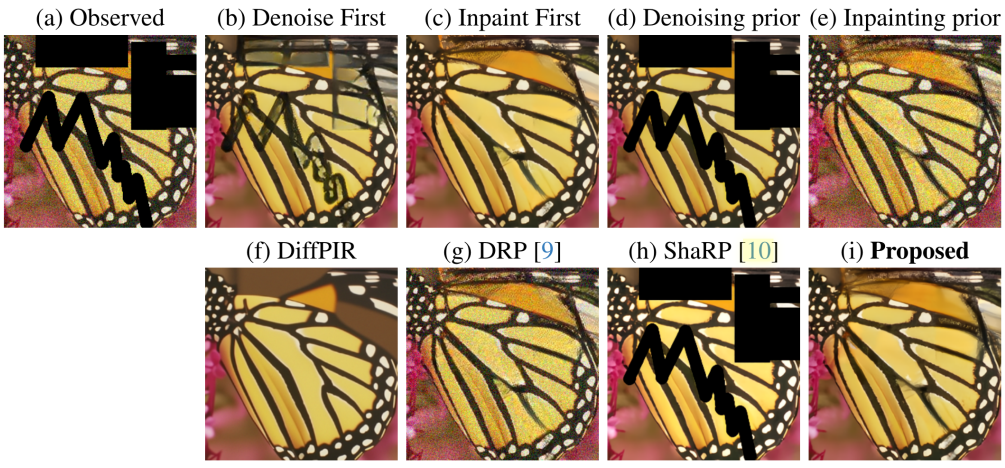


Figure 8: Comparison of different restoration approaches for a noisy inpainting problem with our FiRe algorithm.

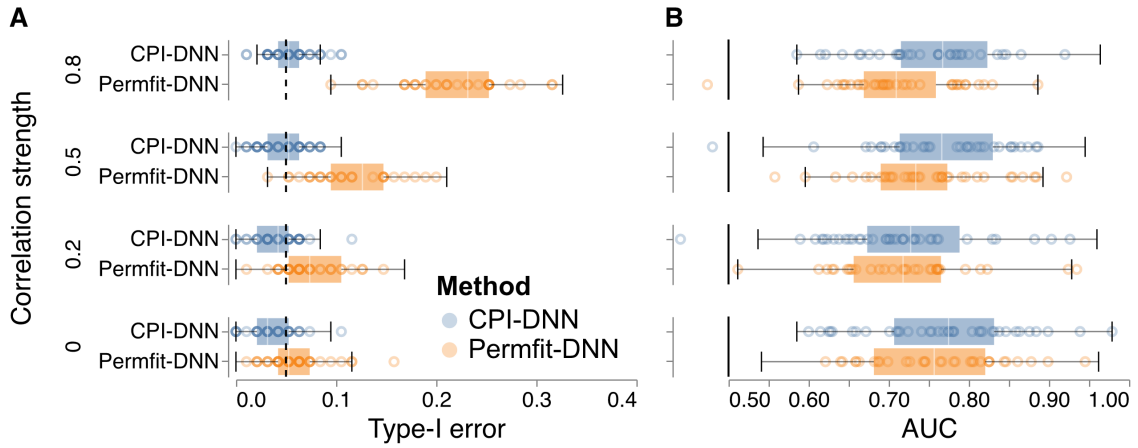


Figure 9: **Performance of Conditional permutation-based vs standard permutation-based variable importance**: Performance at detecting important variables on simulated data with n 300 samples and p 100 features. **(A)**: The type-I error quantifies to which extent the rate of low p-values (p 0.05) exceeds the nominal false positive rate. **(B)**: The AUC score measures to which extent variables are ranked consistently with the ground truth. Dashed line: targeted type-I error rate. Solid line: chance level.

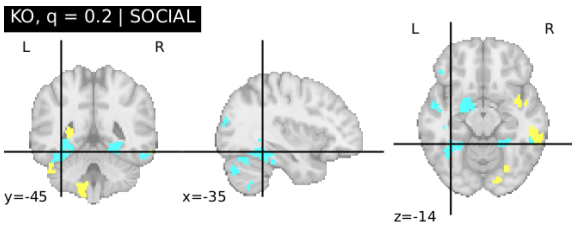


Figure 10: **Application of Kopi to cognitive brain imaging.** We have employed KOPI on fMRI and genomics data. The aim of fMRI data analysis is to recover relevant brain regions for a given cognitive task as shown below. Here we display brain regions whose activity predicts that the participant is atending to stimuli with social motion.

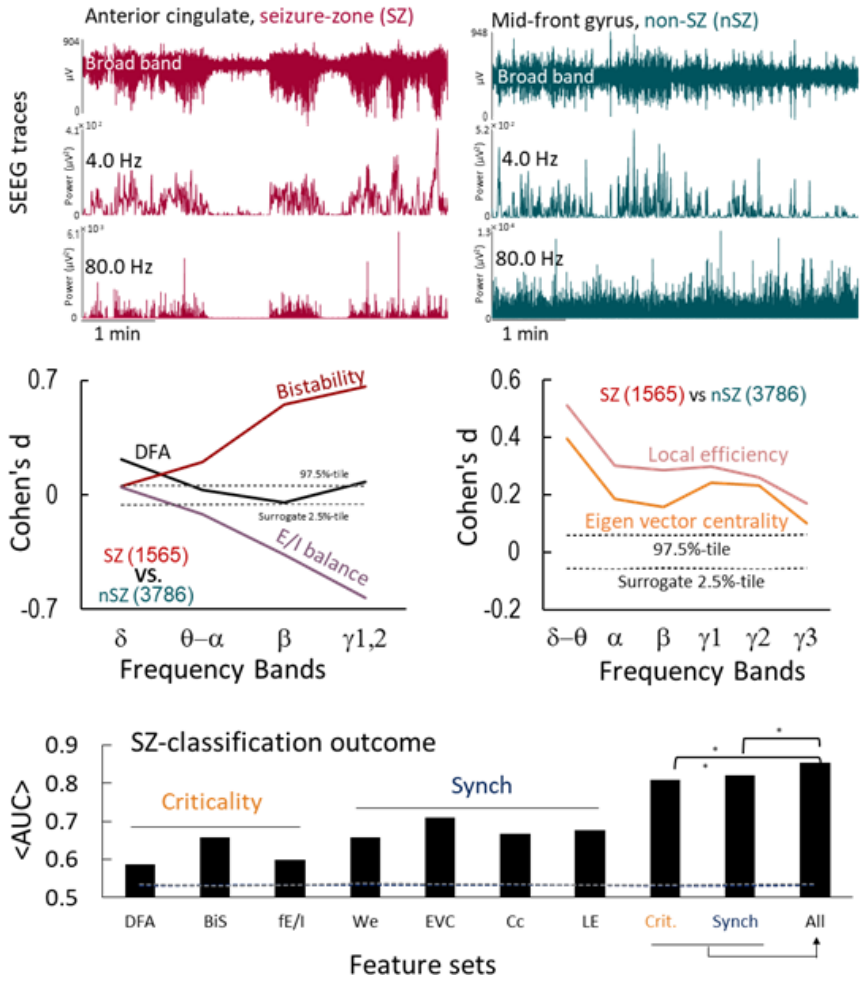


Figure 11: **Individual level evidence of differences between seizure zone (SZ) and non-SZ (nSZ)**. Top: Five minutes of broadband and narrowband traces from (A) an SZ contact and (B) an nSZ contact from the frontal region of a representative subject. Center: Criticality and synchrony assessments differentiated seizure zone (SZ) and non-SZ (nSZ) on the population level using band-collapsed criticality indices. Bottom: Achieving optimal seizure zone (SZ) classification by combining all criticality and synchrony features.