



**HAL**  
open science

## Few-Shot Decoding of Brain Activation Maps

Myriam Bontonou, Giulia Lioi, Nicolas Farrugia, Vincent Gripon

► **To cite this version:**

Myriam Bontonou, Giulia Lioi, Nicolas Farrugia, Vincent Gripon. Few-Shot Decoding of Brain Activation Maps. EUSIPCO 2021: 29th European Signal Processing Conference, Aug 2021, Dublin, Ireland. pp.1326-1330, 10.23919/EUSIPCO54536.2021.9616158 . hal-03675137

**HAL Id: hal-03675137**

**<https://hal.science/hal-03675137v1>**

Submitted on 27 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Few-shot Decoding of Brain Activation Maps

Myriam Bontonou, Giulia Lioi, Nicolas Farrugia, Vincent Gripon

Lab-STICC, IMT Atlantique

Brest, France

firstname.surname@imt-atlantique.fr

**Abstract**—Few-shot learning addresses problems for which a limited number of training examples are available. So far, the field has been mostly driven by applications in computer vision. Here, we are interested in adapting recently introduced few-shot methods to solve problems dealing with neuroimaging data, a promising application field. To this end, we create a neuroimaging benchmark dataset for few-shot learning and compare multiple learning paradigms, including meta-learning, as well as various backbone networks. Our experiments show that few-shot methods are able to efficiently decode brain signals using few examples, which paves the way for a number of applications in clinical and cognitive neuroscience, such as identifying biomarkers from brain scans or understanding the generalization of brain representations across a wide range of cognitive tasks.

**Index Terms**—few-shot learning, classification, neuroimaging data, fMRI, brain decoding

## I. INTRODUCTION

Machine learning, in particular deep learning, has achieved unprecedented performance in a variety of tasks in the past few years. As a matter of fact, when training data is abundant, deep neural networks (DNNs) are often able to extract relevant hidden features to generalize decisions to new data. The need for a large amount of training data is easily explained by the fact that DNNs contain a very large number of trainable parameters, which are required to achieve the best performance.

However, in some domains, data labeling or acquisition can be expensive, thus limiting the potential of deep learning-based methods. In particular, in the case of neuroscience, studies require a massive investment in material and human resources. In this article alone, we use neuroimaging data from a study where 13 people have spent several hours in MRI scanners. And yet, we are far from the amount of data usually required to train a DNN. In an effort to tackle this limitation, the field of few-shot learning has proposed a lot of methods in the past few years (c.f. Section II). A vast majority of these works focus on the case of computer vision, with the double interest of the relative simplicity of the considered signals and the existence of standardized benchmarks.

Here, our primary aim is to stress the ability of recently introduced few-shot methods to solve tasks in the domain of neuroimaging. To this end we compare several few-shot methods on neuroimaging data coming from the collection of the IBC dataset (releases 1 [1] and 2 [2]) on NeuroVault [3]. Our contribution is twofold:

- we create and introduce a neuroimaging classification benchmark for few-shot learning,

- we adapt several vision-based few-shot methods and compare their performance.

The benchmark and our PyTorch code are released publicly on [https://github.com/mbonto/fewshot\\_neuroimaging\\_classification](https://github.com/mbonto/fewshot_neuroimaging_classification).

After carefully selecting the data from the IBC collection, the new benchmark ends up with 106 classes and about 20 examples per class. The classes are conditions linked to tasks performed by subjects during a fMRI scan (e.g. audio sentence, speech sound, face gender, reward, left hand...). The examples are elementary contrast maps derived from the fMRI scans. As these inputs are naturally supported on brain graphs, this benchmark can also be used to test graph-based methods.

The structure of the paper is as follows. In Section II, we introduce related work in the domains of signal processing for neuroimaging and few-shot classification. In Section III, we present the methodology, including the creation of the neuroimaging benchmark for few-shot classification, together with the main backbone networks and methods we consider in this work. In Section IV, we conduct experiments aiming at: 1) stressing the ability of few-shot methods to achieve a better accuracy than a direct classifier trained with a limited amount of data, and 2) comparing the architectures of backbones in terms of signal processing. Finally, Section V is a conclusion.

To the best of our knowledge, few-shot learning has never been applied to neuroimaging, a field where modern deep learning approaches have been applied to the analysis of large datasets only recently [4]–[6]. This work makes an important step towards the application of deep learning to decode brain signals where generally few annotated examples are available. As it aims at fully exploiting the predictive power of DNNs for neuroimaging data, it may impact a number of applications (e.g. identifying biomarkers from brain scans, cognitive neuroscience...).

## II. RELATED WORK

Functional MRI (fMRI) measures brain activity changes caused by the deoxygenation of hemoglobin following neuronal activation (the so-called blood oxygen level dependent, or BOLD signal) [7], paving the way for human brain mapping. While most classical approaches relied on simple linear statistical models of individual signals considered separately [8], machine learning is increasingly being applied to decode brain activity [9], by performing supervised learning of multivariate activation patterns elicited by various tasks [10]–[12]. As brain signals are naturally supported by the connectivity structure of the brain, recent studies suggest using graph

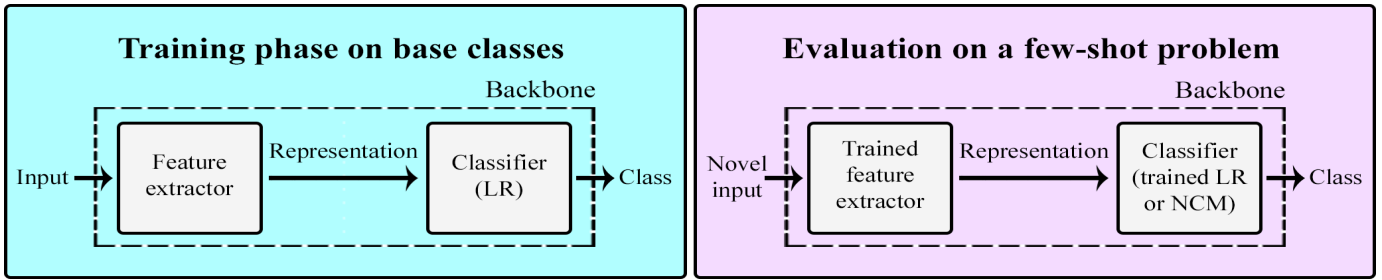


Fig. 1. **Overview of the few-shot learning approach.** During the training phase, the backbone learns rich representations from the abundant examples of base classes. During the evaluation phase, the backbone is adapted to new classes using only a few examples.

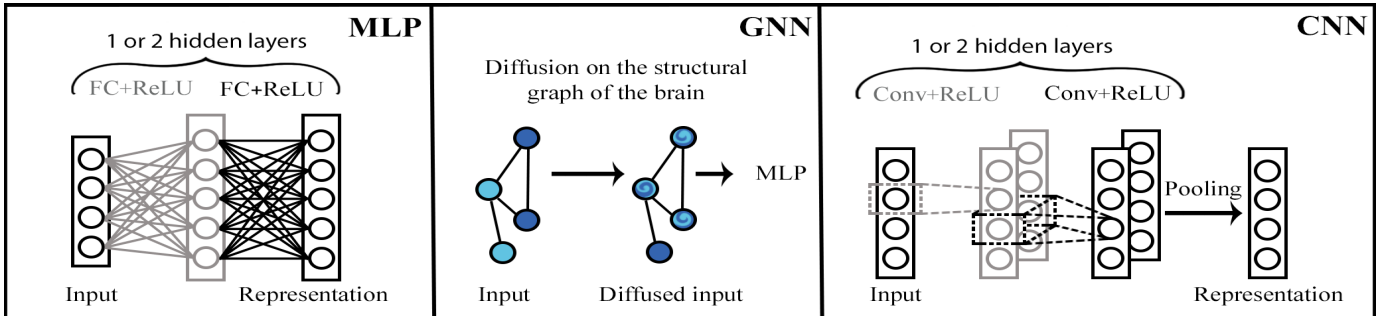


Fig. 2. **Architectures of the feature extractors.** FC stands for fully connected layer and Conv for convolutional layer with  $1 \times 1$  kernels.

neural networks (GNN) to decode brain activity by taking into account brain connectivity [13]–[15]. In addition, a few recent studies have considered transfer learning scenarios; a large dataset is used to train a first classifier, and a smaller dataset exploits the representations learned by the first classifier in a fixed downstream task [6], [16]. Here, we consider the more specific setting of few-shot learning where the downstream task contains very few annotated elements.

The general idea of few-shot learning consists in training a DNN (called *backbone* network) on a lot of data, and then, to transfer it as a prior to solve few-shot problems. The 3 main distinct strategies in the field are the following.

*Representation-based* paradigm [17]–[22]: the first layers of the backbone network are seen as a feature extractor, and the last layer as a simple classifier (e.g. logistic regression). The strategy is to learn, on many classes, how to extract relevant hidden features. When a few-shot problem comes, the features of the few data examples are extracted with the feature extractor. A classifier, containing few parameters, is then trained on these features to recognize new classes.

*Initialization-based* paradigm [23]–[25] (also called meta-learning): the backbone network is trained on many few-shot problems. A problem consists in a set of a few training examples and a set of target samples we aim at classifying. In a first loop, the backbone network trains on the training examples in a few epochs. In a second loop, the initial parameters of the backbone network are updated to minimize its loss evaluated on the target samples. Thus, on a novel few-shot problem, training the backbone in a few epochs should lead to a good performance.

*Hallucination-based* paradigm [26], [27]: the number of training examples is artificially augmented, either by deforming the original examples, or by combining them.

In this work, we focus on the comparison between representation-based and initialization-based, as the metrics used to evaluate hallucination-based methods may be difficult to setup objectively for brain signals.

### III. METHODOLOGY

In this section, we present the standard few-shot classification setting on images. We introduce the few-shot methods we compare in the experiments. We also detail how the neuroimaging dataset is constructed and adapted to the few-shot setting.

#### A. Few-shot setting

In few-shot learning, we use three datasets. The first one, called *base dataset*, contains a large number of labeled data samples corresponding to many classes. The second one, called *validation dataset*, is made of distinct classes. It is meant to evaluate the ability of a DNN trained on the base dataset to generalize to new classes. The final dataset, called *novel dataset*, contains a few examples from novel classes distinct from both the base and the validation datasets. The idea is to exploit the knowledge acquired while training a network on the base dataset in order to circumvent the lack of data in the novel dataset. The network trained on the base dataset is called the *backbone* network.

In the literature, few-shot tasks are often defined in the novel dataset as follows: we consider  $N$ -way (number of

classes),  $K$ -shot (number of labeled examples per class),  $Q$ -query (number of samples to label per class) tasks. The aim is to correctly classify the  $NQ$  unlabeled samples. Two types of methods can be implemented: *inductive few-shot* where the  $NQ$  unlabeled samples are classified independently, and *transductive few-shot* where their joint distribution can be taken into account while performing classification.

### B. Learning paradigms

We propose to evaluate the efficiency of three methods: SimpleShot [17], PT+MAP [21] and MAML++ [24]. SimpleShot and PT+MAP aim at learning good representations for decoding images. They follow the representation-based paradigm. A backbone network is trained to extract generic features enabling to recognize all base classes at once, with the hope that these features will help to classify novel classes. As for MAML++, it aims at learning a good initialization so that the backbone network can easily adapt to the novel dataset with only a few epochs of training, and as such it follows the initialization-based paradigm. SimpleShot and MAML++ are inductive, whereas PT+MAP is transductive. The purpose of PT+MAP is to adjust the distribution of each class in the representation space to fit a Gaussian-like distribution. Further details are in the original papers [17], [21], [24].

### C. Backbone networks

The backbones, made up of a feature extractor and a logistic regression (LR), are trained on the base dataset. In the following, we call *representation* the set of features obtained at the output of the feature extractor. When a new few-shot task comes, the backbones are adapted to new classes using only a few training examples. In MAML++, the whole backbone is fine-tuned for a few epochs. In SimpleShot, a nearest-class mean classifier (NCM) is applied on the representations of the inputs without additional training. Similarly, in PT+MAP, only a new classifier is trained. See Fig. 1 for an illustration.

As mentioned earlier, recent studies advocate the use of GNN to classify neuroimaging data. Here, we try to exploit this graph data structure in the scope of few-shot classification of brain signals. We therefore consider three backbone architectures, described in the following and in Fig. 2.

We first use a multi-layer perceptron (MLP) where the input is seen as a simple vector describing the activity of the considered regions of interest. We vary 2 hyperparameters: the number of hidden layers and the number of features per hidden layer. We then consider a simplified GNN based on the simple graph convolution model [28], where the input is the vector described before. It is diffused once on a structural graph describing the interaction between regions of interest, and then it is handled by a MLP. We vary the same hyperparameters as for MLP. Finally, we use a 1x1 convolutional neural network (CNN), considering each region of interest independently. Before the classifier, the feature maps are averaged per input region. We vary the number of hidden layers and the number of feature maps per hidden layer. For all backbones, as the size of our inputs are small (360), we only consider shallow

neural networks with one or two hidden layers. We vary the number of feature (maps) between 64 and 1024. More details about the architectures and the graph are given on our GitHub.

### D. Neuroimaging benchmark for few-shot learning

The benchmark dataset contains 106 classes which are randomly split into 64 base classes, 21 validation classes and 21 novel classes. Each class has between 21 and 78 samples (median: 33) of size  $105 \times 127 \times 105$ . Here, we consider a summarized version of the samples into 360 regions of interest (ROI). The proposed few-shot-ready dataset is publicly available on the GitHub of the project. In the following, we give additional details explaining the use of the Individual Brain Charting (IBC) dataset [1], the meaning of the classes and the content of the samples.

In the IBC dataset (release 1 and 2), 13 subjects perform 25 tasks covering many aspects of cognition, such as motor movements, emotions, language tasks, working memory, video watching or social tasks. Brain activity is measured simultaneously using functional MRI. Each task enables to study several conditions associated with a wide range of cognitive states. E.g. during the task called *hcp\_motor*, the studied conditions are movement of the left hand, right hand, left foot, right foot and tongue. In total, we included 106 conditions as classes in our benchmark. The full list is available on our GitHub. This variety of experiments makes IBC an interesting benchmark for few-shot learning in decoding cognitive states, as it enables to test whether the variability in a set of cognitive states can be exploited to decode new cognitive states with few examples.

In this benchmark, the data samples are not time-varying 3D measurements of brain activity. Instead, we use the *contrast maps* available on NeuroVault [3] (collection 6618). A contrast map is a statistical map of the effect of a condition on the brain activity of a subject. It is derived from a general linear model (GLM). Broadly speaking, the scans of each participant are aligned to each other and normalized into the standard MNI152 space. Then, they are analyzed using the GLM to capture variations in BOLD signal coupled to neuronal activity pertaining to task performance. This results in a set of weights across the whole brain, therefore considered as a contrast map of the effect of the condition on brain activity with respect to a baseline state. We have at least 21 samples of such maps, corresponding to multiple subjects and repetitions of the experiments.

Finally, in this work, we consider a summarized version of these contrast maps in regions of interest (ROIs). We perform a parcellation of these maps using the 360 ROI of the Glasser atlas [29], a recently proposed atlas of the human brain that integrates multimodal measures of the brain. For the GNN, we use a thresholded structural graph (1% highest connection weights) defined on the same ROI, previously introduced in [30] and used in [31].

## IV. EXPERIMENTS

Following the trend in few-shot learning, we consider the two most common case studies: 5-way 5-shot 15-query and

TABLE I  
AVERAGE ACCURACY AND 95% CONFIDENCE INTERVAL OVER 10000 5-WAY TASKS FROM THE NOVEL DATASET.

Method (setting)	Archi	layers / features	Best backbones		Backbones with representations of size 360		
			5-shot	1-shot	layers / features	5-shot	1-shot
Baseline	-	-	70.56 ± 0.21	57.26 ± 0.20	-	70.56 ± 0.21	57.26 ± 0.20
SimpleShot (induct.) [17]	MLP	2/360	<b>86.00 ± 0.16</b>	<b>72.54 ± 0.20</b>	2/360	<b>86.00 ± 0.16</b>	<b>72.54 ± 0.20</b>
	GNN	2/1024	85.14 ± 0.16	71.96 ± 0.21	2/360	84.96 ± 0.16	71.67 ± 0.20
	CNN	2/64	74.26 ± 0.20	59.98 ± 0.20	2/64	74.26 ± 0.20	59.98 ± 0.20
PT+MAP (trans.) [21]	MLP	2/360	<b>88.76 ± 0.17</b>	<b>84.34 ± 0.25</b>	2/360	<b>88.76 ± 0.17</b>	<b>84.34 ± 0.25</b>
	GNN	2/1024	87.86 ± 0.18	83.19 ± 0.25	2/360	87.02 ± 0.18	82.39 ± 0.25
	CNN	2/64	74.82 ± 0.20	63.71 ± 0.26	2/64	74.82 ± 0.20	63.71 ± 0.26
MAML++ (induct.) [24]	MLP	1/360	82.31 ± 0.18	67.64 ± 0.22	1/360	82.31 ± 0.18	67.64 ± 0.22
	GNN	1/128	80.87 ± 0.18	67.71 ± 0.22	1/360	81.06 ± 0.17	66.47 ± 0.22
	CNN	2/360	76.80 ± 0.22	64.68 ± 0.21	2/360	76.80 ± 0.22	64.68 ± 0.21

5-way 1-shot 15-query tasks. In practice, other settings could be studied depending on the application at stake. The tasks are synthesized by randomly sampling classes and samples in these classes. To obtain statistically robust results on the novel dataset, we average the accuracy over 10000 tasks and we report the confidence interval at 95%. As several hyperparameters have to be defined, we select them using the validation dataset. The ones obtaining the best accuracy on 500 5-way 5-shot tasks generated from this set are kept. As stated in Section III, the backbone networks are trained on the base dataset either to recognize all base classes at once (SimpleShot) or to directly classify few-shot problems (MAML++). When relevant, we use PyTorch’s weighted random sampler to handle class imbalance.

#### A. Comparison of few-shot learning paradigms

We first compare the learning paradigms, while varying the backbone networks and their hyperparameters (number of layers, number of features per layer (MLP and GNN), number of feature maps (CNN)). The results are presented in Table I (best backbones). We also compare the paradigms with a NCM directly applied on the raw data (baseline). Our first observation is that, in the inductive setting, SimpleShot outperforms MAML++. Interestingly, this is similar to experiments with standardized vision benchmarks in the field [17]. All considered paradigms allow significant increase in accuracy when compared to using the raw features, showing the ability of few-shot vision methods to perform well in the considered case of neuroimaging data. Finally, the best results are obtained using the transductive PT+MAP method, even though the gain is less significant when considering 5-shot tasks. As PT+MAP uses additional information from the query samples, it is not surprising.

#### B. Comparison of backbones

We also look at the influence of the backbones and their hyperparameters on each considered learning paradigm. In Table I, the results with GNN are worse than the ones with MLP. GNN fails at better leveraging the underlying structure of this dataset whereas it is naturally supported by a graph. This may be due to the way GNN exploits the graph data

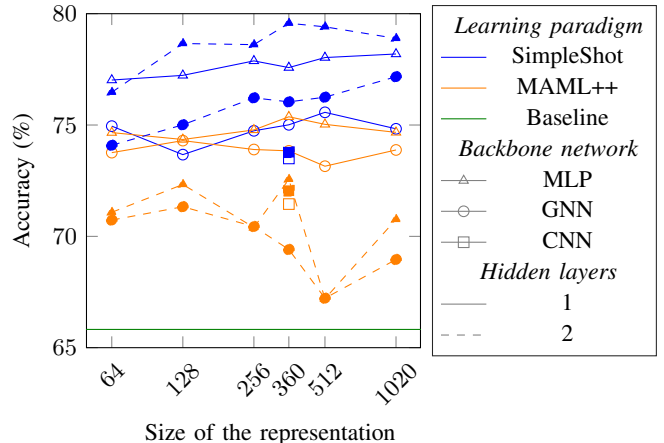


Fig. 3. Average accuracy on 500 5-way 5-shot tasks from the validation split as a function of various hyperparameters. These results are used to select the architecture of the best backbone for the evaluation on novel classes. As PT+MAP uses the same backbone as SimpleShot, it is not represented here.

structure. Here, GNN diffuses the input on a brain graph. Intuitively, it amounts to averaging the value of the input on each ROI with the values of the regions connected to it on the graph. In some applications, this diffusion acts as a denoising process. Here, the brain signals in some tasks might be strongly localized (e.g. motor tasks) while others engage a distributed set of distant brain regions [32].

In Fig. 3, we report the average accuracy obtained on the validation tasks as a function of the size of the representation of the inputs while varying the few-shot paradigms, the backbones and the number of hidden layers. As a reminder, what we call representation of an input is the set of features obtained at the output of the feature extractor. Varying the hyperparameters of each architecture on the validation split enables to select the best hyperparameters for the experiments on the novel classes. As PT+MAP uses the same trained backbone as SimpleShot, it is not represented here. In Fig. 3, we observe that the number of hidden layers and the size of the representation seem to have little influence on overall performance, even though the trend is towards larger dimensions for SimpleShot and smaller for MAML++.

### C. Comparison of learning paradigms for a fixed size of representation

As shown in the previous experiment, the best representations are not necessarily of the same dimension for various backbones. To provide a fair comparison, we report, in Table I (backbones with representations of size 360), the performance obtained on the novel dataset when all representations are of size 360. Note that we chose 360 because all CNN feature extractors output representations of this dimension. In our case, there are new results only for GNN. We observe that previous conclusions are still valid.

### D. Influence of the split

Finally, we perform experiments with another random split of the classes into training/validation/novel datasets to assess that few-shot learning methods still perform better than the simple baseline. The detailed results are given on our GitHub repository. We get the same conclusions.

## V. CONCLUSION

In this paper, we introduced a neuroimaging benchmark dataset for few-shot learning. By comparing various few-shot methods using several backbone architectures, we achieved engaging results: few-shot vision methods perform well in the context of neuroimaging data. However, the backbone architectures we tested had little influence on the results. In perspective, we will investigate how to conceive a backbone that better leverages the graph data structure of this dataset.

This work opens up a pathway towards the application of deep learning to decode brain signals, where generally few examples are available. We plan to extend this study to 3D time-varying fMRI brain signals. In future work, it would also be interesting to evaluate the few-shot performance over cross-domain datasets.

## REFERENCES

- [1] A. L. Pinho, A. Amadon, T. Ruest, M. Fabre, E. Dohmatob, I. Denghien, C. Ginisty, Séverine-Becuwe, S. Roger, L. Laurier, V. Joly-Testault, G. Médiouni-Cloarec, C. Doublé, B. Martins, P. Pinel, E. Eger, G. Varoquaux, C. Pallier, S. Dehaene, L. Hertz-Pannier, and B. Thirion, "Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping," *Sci Data*, vol. 5, p. 180105, 2018.
- [2] A. L. Pinho, A. Amadon, B. Gauthier, N. Clairis, A. Knops, S. Genon, E. Dohmatob, J. J. Torre, C. Ginisty, S. Becuwe-Desmidt *et al.*, "Individual brain charting dataset extension, second release of high-resolution fmri data for cognitive mapping," *Sci Data*, vol. 7, no. 1, pp. 1–16, 2020.
- [3] 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," *Front. Neuroinform.*, vol. 9, p. 8, 2015, link: <https://identifiers.org/neurovault.collection:6618>.
- [4] M.-A. Schulz, B. T. Yeo, J. T. Vogelstein, J. Mourao-Miranada, J. N. Kather, K. Kording, B. Richards, and D. Bzdok, "Different scaling of linear models and deep learning in ukbiobank brain images versus machine-learning datasets," *Nature Commun.*, vol. 11, no. 1, pp. 1–15, 2020.
- [5] A. Abrol, Z. Fu, M. Salman, R. Silva, Y. Du, S. Plis, and V. Calhoun, "Hype versus hope: Deep learning encodes more predictive and robust brain imaging representations than standard machine learning," *BioRxiv*, 2020.
- [6] T. He, L. An, J. Feng, D. Bzdok, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo, "Meta-matching: a simple framework to translate phenotypic predictive models from big to small data," *BioRxiv*, 2020.
- [7] R. B. Buxton, *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge University Press, 2009.
- [8] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *HBM*, vol. 2, no. 4, pp. 189–210, 1994.
- [9] G. Varoquaux and R. A. Poldrack, "Predictive models avoid excessive reductionism in cognitive neuroimaging," *Current Opinion Neurobio.*, vol. 55, pp. 1–6, 2019.
- [10] N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-based functional brain mapping," *Proc. Nat. Academy of Sci.*, vol. 103, no. 10, pp. 3863–3868, 2006.
- [11] N. Kriegeskorte, "Pattern-information analysis: from stimulus decoding to computational-model testing," *NeuroImage*, vol. 56, no. 2, pp. 411–421, 2011.
- [12] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nature Reviews Neurosci.*, vol. 7, no. 7, pp. 523–534, 2006.
- [13] Y. Zhang, L. Tetrel, B. Thirion, and P. Bellec, "Functional annotation of human cognitive states using deep graph convolution," *NeuroImage*, vol. 231, p. 117847, 2021.
- [14] A. Kazeminejad and R. C. Sotero, "Topological properties of resting-state fmri functional networks improve machine learning-based autism classification," *Front. Neurosci.*, vol. 12, p. 1018, 2019.
- [15] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Distance metric learning using graph convolutional networks: Application to functional brain networks," in *Int. Conf. MICCAI*. Springer, 2017, pp. 469–477.
- [16] Y. Zhang and P. Bellec, "Transferability of brain decoding using graph convolutional networks," *BioRxiv*, 2020.
- [17] Y. Wang, W.-L. Chao, K. Q. Weinberger, and L. van der Maaten, "Simple-shot: Revisiting nearest-neighbor classification for few-shot learning," *arXiv:1911.04623*, 2019.
- [18] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, and J.-B. Huang, "A closer look at few-shot classification," in *ICLR*, 2018.
- [19] P. Mangla, N. Kumari, A. Sinha, M. Singh, B. Krishnamurthy, and V. N. Balasubramanian, "Charting the right manifold: Manifold mixup for few-shot learning," in *IEEE WACV*, 2020, pp. 2218–2227.
- [20] M. Lichtenstein, P. Sattigeri, R. Feris, R. Giryes, and L. Karlinsky, "Tafssl: Task-adaptive feature sub-space learning for few-shot classification," *arXiv:2003.06670*, 2020.
- [21] Y. Hu, V. Gripon, and S. Pateux, "Leveraging the feature distribution in transfer-based few-shot learning," *arXiv:2006.03806*, 2020.
- [22] ———, "Exploiting unsupervised inputs for accurate few-shot classification," *arXiv:2001.09849*, 2020.
- [23] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *ICML*, 2017.
- [24] A. Antoniou, H. Edwards, and A. Storkey, "How to train your maml," in *ICLR*, 2018.
- [25] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv:1803.02999*, 2018.
- [26] H. Zhang, J. Zhang, and P. Koniusz, "Few-shot learning via saliency-guided hallucination of samples," in *Proc. IEEE Conf. CVPR*, 2019, pp. 2770–2779.
- [27] Z. Chen, Y. Fu, Y.-X. Wang, L. Ma, W. Liu, and M. Hebert, "Image deformation meta-networks for one-shot learning," in *Proc. IEEE Conf. CVPR*, 2019, pp. 8680–8689.
- [28] F. Wu, A. H. Souza Jr, T. Zhang, C. Fifty, T. Yu, and K. Q. Weinberger, "Simplifying graph convolutional networks," in *ICML*, 2019.
- [29] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. V. Essen, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536 (7615), pp. 171–178, 2016.
- [30] M. G. Preti and D. Van De Ville, "Decoupling of brain function from structure reveals regional behavioral specialization in humans," *Nature Commun.*, vol. 10, no. 1, pp. 1–7, 2019.
- [31] A. Brahim and N. Farrugia, "Graph fourier transform of fmri temporal signals based on an averaged structural connectome for the classification of neuroimaging," *Artif. Intell.*, p. 101870, 2020.
- [32] J. R. Cohen and M. D'Esposito, "The segregation and integration of distinct brain networks and their relationship to cognition," *J. Neurosci.*, vol. 36, no. 48, pp. 12 083–12 094, 2016.