RESEARCH ARTICLE

Nonlinear latent representations of high-dimensional task-fMRI data: Unveiling cognitive and behavioral insights in heterogeneous spatial maps

Mariam Zabihi[1,2,3]*, Seyed Mostafa Kia[1,2,4], Thomas Wolfers[1,2,5,6], Stijn de Boer[1], Charlotte Fraza[1,2], Richard Dinga[1,2], Alberto Llera Arenas[1], Danilo Bzdok[7,8], Christian F. Beckmann[1,2,9], Andre Marquand[1,2,10]

1 Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Nijmegen, the Netherlands, **2** Department for Cognitive Neuroscience, Radboud University Medical Center Nijmegen, Nijmegen, the Netherlands, **3** MRC Unit for Lifelong Health & Ageing, University College London (UCL), London, United Kingdom, **4** Department of Psychiatry, University Medical Center Utrecht, Utrecht, the Netherlands, **5** NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway, **6** Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Tübingen, Germany, **7** Multimodal Imaging and Connectome Analysis Lab, McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada, **8** Mila - Quebec Artificial Intelligence Institute, Montreal, Quebec, Canada, **9** Centre for Functional MRI of the Brain, University of Oxford, Oxford, United Kingdom, **10** Department of Neuroimaging, Institute of Psychiatry, Psychology, & Neuroscience, King's College London, London, United Kingdom

* mariam.zabihi@gmail.com

OPEN ACCESS

Citation: Zabihi M, Kia SM, Wolfers T, de Boer S, Fraza C, Dinga R, et al. (2024) Nonlinear latent representations of high-dimensional task-fMRI data: Unveiling cognitive and behavioral insights in heterogeneous spatial maps. PLoS ONE 19(8): e0308329. https://doi.org/10.1371/journal.pone.0308329

Editor: Andrea Tangherloni, Bocconi University: Universita Bocconi, ITALY

Received: October 18, 2023

Accepted: July 22, 2024

Published: August 8, 2024

Data Availability Statement: The GitHub links for the code are: https://github.com/mariam186/semi-supervised-3-D-autoencoder and https://github.com/amarquand/PCNtoolkit. Regarding the data, the UK Biobank and the Human Connectome Project (HCP) are not publicly accessible. They can be accessed by request through universities and research centers. More information can be found at the following links: https://db.humanconnectome.org (HCP) and https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access (UKB).

Abstract

Finding an interpretable and compact representation of complex neuroimaging data is extremely useful for understanding brain behavioral mapping and hence for explaining the biological underpinnings of mental disorders. However, hand-crafted representations, as well as linear transformations, may inadequately capture the considerable variability across individuals. Here, we implemented a data-driven approach using a three-dimensional autoencoder on two large-scale datasets. This approach provides a latent representation of high-dimensional task-fMRI data which can account for demographic characteristics whilst also being readily interpretable both in the latent space learned by the autoencoder and in the original voxel space. This was achieved by addressing a joint optimization problem that simultaneously reconstructs the data and predicts clinical or demographic variables. We then applied normative modeling to the latent variables to define summary statistics ('latent indices') and establish a multivariate mapping to non-imaging measures. Our model, trained with multi-task fMRI data from the Human Connectome Project (HCP) and UK biobank task-fMRI data, demonstrated high performance in age and sex predictions and successfully captured complex behavioral characteristics while preserving individual variability through a latent representation. Our model also performed competitively with respect to various baseline models including several variants of principal components analysis, independent components analysis and classical regions

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

of interest, both in terms of reconstruction accuracy and strength of association with behavioral variables.

Background

An important challenge in the application of machine learning to neuroimaging is to find an optimal low dimensional summary or representation of the complex spatial information encoded in brain images into a biologically interpretable readout that preserves important relationships between data points. These representations can be used to understand inter-individual differences, ascertain association between neuroimaging data and cognitive variables, and identify biomarkers that can be used to understand and make predictions on the basis of the biological underpinnings of both healthy and disordered mental states [1–5].

Deep neural network models are one candidate for providing such a representation. However neuroimaging studies have traditionally had a limited number of high-dimensional datasets, which until recently had hindered the use of complex deep neural network models for a time due to the curse of dimensionality [6]. The recent increase in the availability of large-scale neuroimaging data has provided a great opportunity to move toward using complex nonlinear methods, for example, based on deep learning approaches [7–14]. Many deep learning studies in neuroimaging use hand-crafted features [5, 15–18] e.g., regions of interest (ROIs) or image-derived phenotypes (IDPs)- which are potentially suboptimal for prediction because: first, hand-crafted features may fail to capture complex structural or functional characteristics of the brain. Individual differences are often encoded in intricate and overlapping patterns in the brain that are important for understanding its relationship with behavior. Hand-crafted features might not represent these complexities accurately, leading to inferior predictions. Second, these studies do not benefit from the strength of deep neural networks in automatically learning the optimal representation from the data, achieved when the model converges, such as through the use of convolutional filters. This highlights the need for more effective methods to learn representations of neuroimaging data that can accurately predict clinical and cognitive variables. Particularly in task fMRI studies, which are designed to explore mappings from brain activations to cognition and behavior, numerous challenges exist. These includes the extensive heterogeneity across individuals, finding a comprehensive representation, and the need for a reliable reference to compare the activations [19–25]. Consequently, using hand-crafted features potentially leads to losing relevant information, such as inter-individual differences in functional anatomy [5, 26]. In these scenarios, learning a representation of high-dimensional neuroimaging data—rather than using predefined ROIs, for example—may enable a better understanding of individual variations and lead to more accurate predictions of clinical and cognitive measures. Such latent representations allow us to reduce the data dimensionality and extract only the essential features. In other words, a latent representation maps complex and high-dimensional data into a reduced, low-dimensional space [27]. Thus, our research question is: How can we leverage nonlinear techniques to learn a generic or general-purpose latent representation of task-fMRI (tfMRI) data that is not bound to a specific task and accurately predicts a wide range of cognitive and clinical variables?

Motivated by the limitations of existing approaches and the potential benefits of a general-purpose latent representation, we propose a method to learn such a representation from tfMRI data. Most applications of deep learning in neuroscience focus on learning a latent representation optimized for a single supervised learning problem, such as predicting age or sex (e.g., [7, 11, 28, 29]). However, this approach may reduce the generalizability of the learned latent representation to other problems. Our approach aims to overcome this limitation by learning a general-purpose latent space that is not bound to a specific task but instead captures features from the data that are predictive of wide range of cognitive scores. Numerous efforts have been made towards this end [30–37]. Most of these studies evaluate the data representations on the basis of specific measures like reconstruction error but this does not necessarily suggest that the latent space captures relevant features. Although linear data-driven transformations like Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [38–42] are widely used for feature representation and dimensionality reduction in neuroimaging, these methods often fail to extract complex nonlinear relationships in the data [43, 44] To address this challenge, we introduce a 3-dimensional semi-supervised autoencoder (AE) that learns a nonlinear latent space representation of tfMRI images, capturing relevant features and their associations with nIDPs.

In this paper, we propose exploring the value of learning a general purpose nonlinear latent space representation of tfMRI contrast images using a 3-dimensional semi-supervised AE.

Autoencoder neural networks are powerful tools in various applications in neuroimaging studies, from image segmentation to abnormality detection and latent representation [8, 9, 33, 45–48]. Briefly, an autoencoder is a deep neural network architecture consisting of two parts: an encoder and a decoder. The encoder projects the inputs to a lower-dimensional latent space using a non-linear transformation. The decoder translates back the latent space to the original space by reconstructing the inputs [49]. Complementary to existing approaches, we demonstrate how we can control the learned latent representation by incorporating a supervised learning term into the reconstruction, that is, within a joint optimization framework. In our approach, we tailor the search space by adding age and sex to the loss function minimized by the model, ensuring the learned latent representation is not limited to a specific task. Our approach, in contrast to many previous methods, does not require the prior specification of regions of interest (i.e. which can be used as nodes in the network). Instead, it can learn overlapping representations, use the full range of spatial patterns in the fMRI signal, and leverage the strengths of deep learning, such as learning convolutional filters that capture low-level features of the images.

Having learned the latent representation, we then assess whether it demonstrates a stronger association with cognitive, clinical and demographic variables - collectively referred to as ‘non-imaging-derived phenotypes’ (nIDPs) - compared to data in the original space (e.g., mapping from raw image data or hand-crafted features to behavioral scores) or to traditional linear dimensionality reduction techniques such as PCA.

More specifically, in a fully data-driven approach shown in Fig 1, we showed that there is useful information about the data in the nonlinear latent space that is not fully captured by a linear data representation and that such information can be extracted using a hierarchical non-linear autoencoder architecture with joint optimization for age and sex prediction [5, 8]. Here, we employed an autoencoder with an architecture designed from the ground up for tfMRI data and provided a method for visualizing, exploring, and interpreting the learned representation. Last, in this study, we aimed to illustrate how our model can be employed to understand inter-individual differences in brain activity. To achieve this, we applied a normative model [50–52] to a compressed representation of the latent variables, which was derived using the Uniform Manifold Approximation and Projection (UMAP) technique. This approach allowed us to separate age-related variation from other sources of variability, thereby providing a more detailed insight into the factors influencing brain function. We use these deviations for detecting associations with nIDPs. We first trained our model with multi-task fMRI data derived from the Human Connectome Project (HCP) [19] which provides whole-

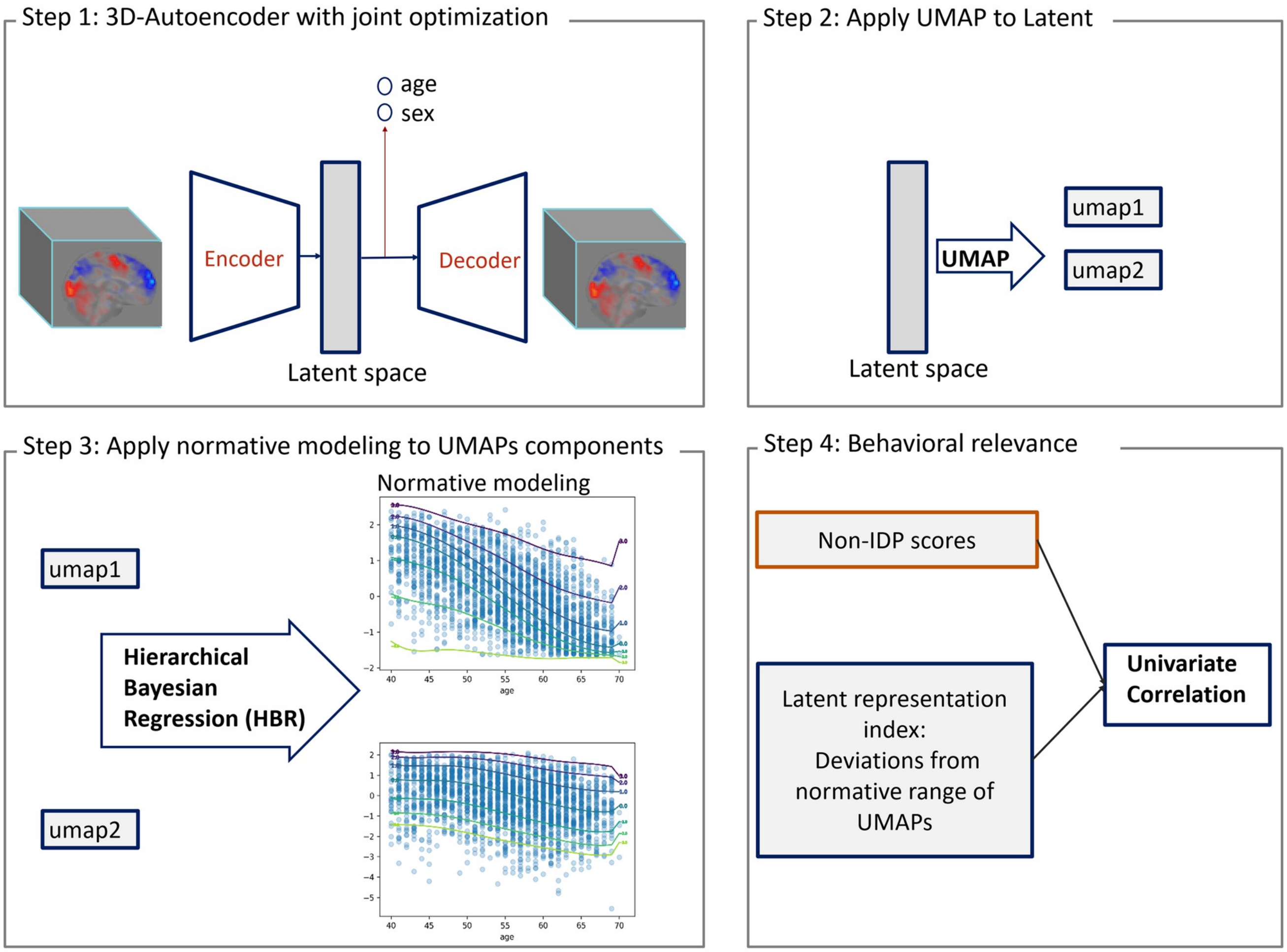
Step 1: 3D-Autoencoder with joint optimization
○ age
○ sex
Encoder
Decoder
Latent space
Step 2: Apply UMAP to Latent
UMAP
umap1
umap2
Latent space
Step 3: Apply normative modeling to UMAPs components
Normative modeling
umap1
Hierarchical Bayesian Regression (HBR)
umap2
age
Step 4: Behavioral relevance
Non-IDP scores
Latent representation index: Deviations from normative range of UMAPs
Univariate Correlation

Fig 1. Method overview: step1) training semi-supervised AE model with joint optimization of age and sex prediction. Step2) applying UMAP transformation to the latent variables of semi-supervised AE. step3) applying HBR normative modeling to the components of UMAPs. Step4) measuring the correlation of non-imaging scores (behavioral, cognitive and clinical scores) and the deviation value from normative range of UMAP components (latent representation index).

https://doi.org/10.1371/journal.pone.0308329.g001

brain coverage across a range of cognitive tasks. We then fine-tuned the network using UK Biobank dataset [23]. Our experimental results show that our nonlinear data representation provides a strong foundation for subsequent analysis of brain-behavior mappings and results in strong associations between our latent index and unseen nIDPs.

Methods

Data

Two different datasets were used in this study. This first dataset consists of tfMRI data from the HCP [19] S500 release. The second tfMRI dataset is from the 2020 UK Biobank imaging release [53].

HCP. We used tfMRI contrast data from 468 participants in total (187 males and 281 females, Age = 29.2±3.5) from seven different behavioral and cognitive tasks (emotion processing, gambling, language, relational processing, social cognition, motor, working memory) across 86 contrasts (86 sub-tasks). These tasks served as the basis in previous brain-imaging work [54, 55]. This yields a total of N≈40K tfMRI scans. The HCP dataset is well suited for this purpose because the task battery covers a wide range of cognitive domains and the neuronal activations associated with the task provide good coverage of the entire brain [20]. The number of participants may vary from task to task; not all the participants have data in all the tasks. While HCP has a large number of samples, the number of participants is relatively small. Therefore, we split data into 5 subsets in a 5-fold cross-validation scheme. The splits are made at the subject level, ensuring that each fold contains all the contrasts for a specified set of subjects. This prevents overly optimistic estimates of generalizability due to the correlations between different contrasts from the same subject. More specifically, in each fold, about 95 participants (20% of the data) were reserved for the test set (N = 8K brain scans) and the rest for the training (N = 32K brain scans, 373 participants). For each fold, we trained a separate model. Moreover, to further guard against overfitting, an independent set of subjects were used to determine the model architecture and optimize hyperparameters (see below and in the S1 File for details).

UK Biobank. We used UK Biobank tfMRI contrast data from 20,781 participants and 5 contrasts, in total N≈104K scans (9,860 males, 10,921 females, Age = 54.6 ±7.4). The tfMRI data derived from UK Biobank uses the same paradigm as the emotion task from the HCP with only minor modifications (e.g. to accommodate shorter run length) [23, 53]. We randomly selected N = 15585 of participants for the train set and 5196 for the test set. All the contrast-models employ the same dataset configuration (the test and train sets).

Non-imaging-derived phenotypes (nIDP) data. The UK Biobank study provides an extensive number of clinical, behavioral, lifestyle and cognitive scores, which we categorized to seven groups e.g., cognitive phenotypes, lifestyle, and mental health (see S1 File for the full list of categories). We only included the measures that their scores are available more than half of participants. Moreover, in line with previous studies [56, 57] the measures that had same value for more than 80% of the participants were excluded from further analysis.

Image preprocessing

For both datasets we used the volumetrically preprocessed images in standard reference space provided by the respective consortia [58, 59] (for HCP using the 'minimally processed' pipeline [58]). Subsequently the scans were downsampled from 2mm to 3mm voxel resolution to reduce the computational burden then cropped tightly to the whole brain such that the dimension of the image decreased to 56×64×56. The model was trained on the whole-brain contrast images.

Model architecture

We developed a deep 3D-convolutional autoencoder that learns to encode and decode tfMRI images using HCP data. Since many choices need to be made regarding the architecture of the autoencoder, we performed a pilot study on a subset of data that was discarded before fitting the final model. Here, we selected the architecture for the autoencoder using held-out data (N = 30 participants reserved data, 2580 scans). Full details of this procedure are provided in the S1 File. The final architecture was as follows: Each encoder and decoder of the semi-supervised AE had three hidden convolutional layers with 3x3x3 kernel size. The bottleneck of the model is a dense layer containing 100 nodes. Each layer, except the output layer, was followed by ReLU activation function [60] to add non-linearity and sparsity to the network and to reduce the likelihood of vanishing gradient. The output layer was followed by a linear activation function. To increase the robustness of the model and avoid overfitting, we incorporated drop-out [61] (drop-out level = 0.2) in each layer except the output layer. To avoid the risk of a degenerate solution, where the autoencoder simply learns the identity function, we added

Gaussian noise [62] (mean = 0, standard deviation = 0.1) to the input layer to randomly corrupt the data (see S1 File for details about the optimization of the architecture of the semi-supervised AE).

The loss function to train the model contains two parts; an unsupervised and a supervised loss. The unsupervised loss simply is the mean squared error of the reconstruction image of the noisy image and the original image. The supervised loss is incorporated in order to control of latent space of the autoencoder; Here, we added age and sex as the supervised part of calculating the loss function. Without this, we cannot guarantee that the learned representation would contain any information about relevant demographic features. We used age as a continuous variable rather than a one-hot encoded matrix (i.e. which would effectively treat the regression as a classification problem [63]). This enables us to generalize beyond the age range used in the training dataset, which is important for transfer learning because of potential differences between cohorts. So the training loss is defined by:

$$loss = \lambda(x - \hat{x})^2 + (1 - \lambda)\Big(|y_{age} - \hat{y}_{age}| + Binary\ crossentropy(y_{sex} - \hat{y}_{sex})\Big)$$

which x is the input image and y_{age} and y_{sex} are age and sex. The first term refers to unsupervised loss which is the usual autoencoder loss and the second term refers to supervised loss. To balance the supervised and unsupervised loss in terms of scale, we used coefficient λ which specifies the importance of supervised loss e.g., $\lambda = 1$ means a completely unsupervised autoencoder (i.e. 'vanilla AE').

We also trained our model on the held-out calibration dataset with different range of $\lambda = 1$, 0.995, 0.95, 0.5, 0.05, 0.005 to select the optimum value of λ in terms of unsupervised and supervised loss.

Training the model

The training data were normalized to have zero mean unit variance across each feature. The layers weights were initialized using Xavier initialization [64]. First, the model was trained using HCP data with 1,000 epochs and using Adam [65] optimizer with an adaptive learning rate. The base learning rate was set at 0.001 and with exponential learning rate decay over each epoch reached 0.0003. Last, the mini-batch gradient descent was conducted with the size of 10 images.

Having trained the model by HCP, the network was trained again using the same hyperparameters with UKB data. Although, this procedure has a similar motivation to fine-tuning, in fact all the weights were re-estimated (i.e. none of the layers were frozen). We consider this to be the most appropriate approach because the age range is very different across these two datasets. Here, the weights of the trained model by HCP were used as initial weights. The base learning rate was reduced to 0.0003 for training with UKB data to limit significant modifications to the pre-trained weights, thereby preserving the valuable features already learned during the initial training phase.

Visualising the latent space representation using UMAP

To visualize and evaluate our model quantitatively, we visualized the latent space using a UMAP approach [66] with two components. UMAP,a manifold learning technique similar to t-distributed stochastic neighbor embedding (t-SNE) [67], preserves the local structure of high dimensional data in a nonlinear space. UMAP is superior to tSNE since it better preserves the global structure of data, in addition to its local structure. Furthermore, it is more stable under perturbation or resampling of the data.

To visualize the latent space with two UMAP components, a UMAP model was fitted using latent variables derived from the training set. To avoid over-engineering the results, we applied UMAP with the default parameter settings. More specifically, the size of local neighborhood to learn the manifold structure of the data was set to 15 while the minimum distance of each data in the low dimensional representation was 0.1 in Euclidean distance. Later, this model was applied to the predicted latent variables of test images. We leave further optimization of these parameters for future work.

We provide several simple examples to demonstrate how this representation helps us understand the distinctiveness, overlap, and distribution of different stimulus classes at different scales. For example, the derived latent representation facilitates the assessment of functional differentiation, determines whether different tasks have distinct or overlapping distributions, helps decide whether a gradient-based or discrete representation may be more appropriate for the data. Moreover, the semi-supervised autoencoder allows us to probe the degree to which functional differentiation changes as a function of different covariates (e.g. whether different tasks or contrasts are differentially affected by ageing processes).

To illustrate, we begin by showcasing the functional differentiation among all HCP tasks, then focus specifically on the motor task, which has three different experimental conditions, based on whether participants were moving their tongue (T), the left hand (LH), right hand (RH) or left or right foot (LF/RF; see Barch et al 2013 [20] for details on the task paradigm). To achieve this, we first calculate the center of the latent space for each task, representing the average response. We then transfer these centers back to the original space using the decoder to generate the corresponding brain images. This is of interest because the classical notion of a relatively continuous 'homunculus' representation of the motor cortex has recently been challenged suggesting a higher degree of functional differentiation than has previously been appreciated [68].

Associations with nIDPs data

Normative modeling of UMAP. To show how our latent space can be used to provide biomarkers that can be then used to predict external behavioural, clinical, or demographic variables–often referred to as nIDPs–we calculated the linear association between clinical and behavioral measures and the deviations of the UMAP-reduced latent space for UK Biobank data. To establish a baseline for comparison, we additionally calculated the associations of nIDPs with UMAPs derived from PCA (number of components = 100), ICA (number of components = 100), and sub-cortical ROIs. However, since the latent variables are related to age and age has a strong association with many cognitive and behavioral scores, we employed normative modeling on the latent space to separate variation that is principally age-related (encoded by the normative model) from inter-individual differences that manifest as deviations from an expected age-related pattern (encoded in the deviations from the normative model). The normative modeling approach has been used extensively to model heterogeneity in various psychiatric disorders [18, 69, 70]. Briefly, this approach provides a statistical estimation of the distribution of brain measures along with the deviations from the reference cohort at the level of each individual participant.

We define the 'latent index' as a feature that indicates the deviation from the normative UMAP of latent variables of each image. To construct this index, we applied normative modeling using a flexible generalization of hierarchical Bayesian regression (HBR) [71–73] to the UMAP of latent variables to remove both linear and non-linear associations with age and sex. Importantly, we used a recent generalization of the HBR method that can handle heteroskedastic and non-Gaussian distributions. Age was defined as a regressor and sex as batch effects.

(See de Boer 2022 [73] and S1 File for details of HBR normative model). In this way, for each UMP component of each individual, we obtained the deviation or z-score, which we refer to as the 'latent index'. Then, we used the latent index as an indicator of individualized brain activation variability by measuring the associations of the latent index and nIDPs using Spearman correlation. To compare this association with linear approaches, we calculated the association between the normative UMAPs of latent variables derived from PCA (number of components = 100) and nIDPs.

Results

Autoencoder performance

As described above and in detail in the S1 File, the optimal number of nodes for each layer and the number of layers of semi-supervised AE model were obtained by a pilot study using independent data and resulted 32, 16, and 8 nodes respectively for the 3 layers of the encoder and 8, 16, 32 for the decoder, respectively. Lambda (λ) was set empirically at 0.05 in to balance the supervised and unsupervised loss (see S1 File for more details on the architecture of semi-supervised AE and the latent space visualization for different values of lambda). Applying the HCP trained model (without fine-tunning) to UKB data resulted in reconstruction error of 0.31 and 0.17 when $\lambda = 0.05$ and $\lambda = 1$, respectively. To elucidate, increasing λ reduced the reconstruction error in the HCP dataset from 0.26 to 0.23, a decrease of 11%. However, this difference was more pronounced in the UKB dataset, where the error reduction was 45%, indicating that lower values of λ significantly impair the autoencoder's performance not only within the original dataset but also more substantially in a secondary dataset. This shows the latent is not fully generalizable, partially because of age differences between this group in addition to other factors such as demographic differences, differences in signal to noise ratio etc. This motivates retraining the model with UKB data. The out-of-sample performance of the models is shown in Table 1. See S4 Table in S1 File with the extensive comparison with PCA.

Visualization of latent space

The scatterplot of UMAP projections of the autoencoder's latent variables is shown in Fig 2 for selected contrasts (one per task) [20] in the HCP dataset and the Faces-Shapes contrast from the emotion task in UKB. This figure shows how the data points are distributed in the latent space, indicating the distinctiveness (i.e., functional differentiation) of each task with respect to one another and with regard to age and sex. By contrasting the left and right columns of Fig 2A and 2B, its clear that: (i) in the vanilla AE ($\lambda = 1$) age and sex are not clearly reflected in the latent space, and rather the latent space principally reflects differences between different tasks; (ii) in the semi-supervised AE ($\lambda = 0.05$), age and sex are more clearly evident in the latent

Table 1. Model performance.

		HCP	*UKB*
Semi-Supervised Autoencoder **($\lambda = 0.05$)**			
	Image reconstruction error (MSE)	0.26± 0.00	0.16 ± 0.02
	Age mean absolute error(MAE)	3.13 ± 0.09	4.84 ± 0.25
	Sex prediction accuracy	81% ± 3%	89% ± 3%
Vanilla Autoencoder **($\lambda = 1$)**			
	Image reconstruction error (MSE)	0.23 ± 0.00	0.15 ± 0.01
PCA			
	Image reconstruction error (MSE)	0.22 ± 0.00	0.14 ± 0.01

https://doi.org/10.1371/journal.pone.0308329.t001

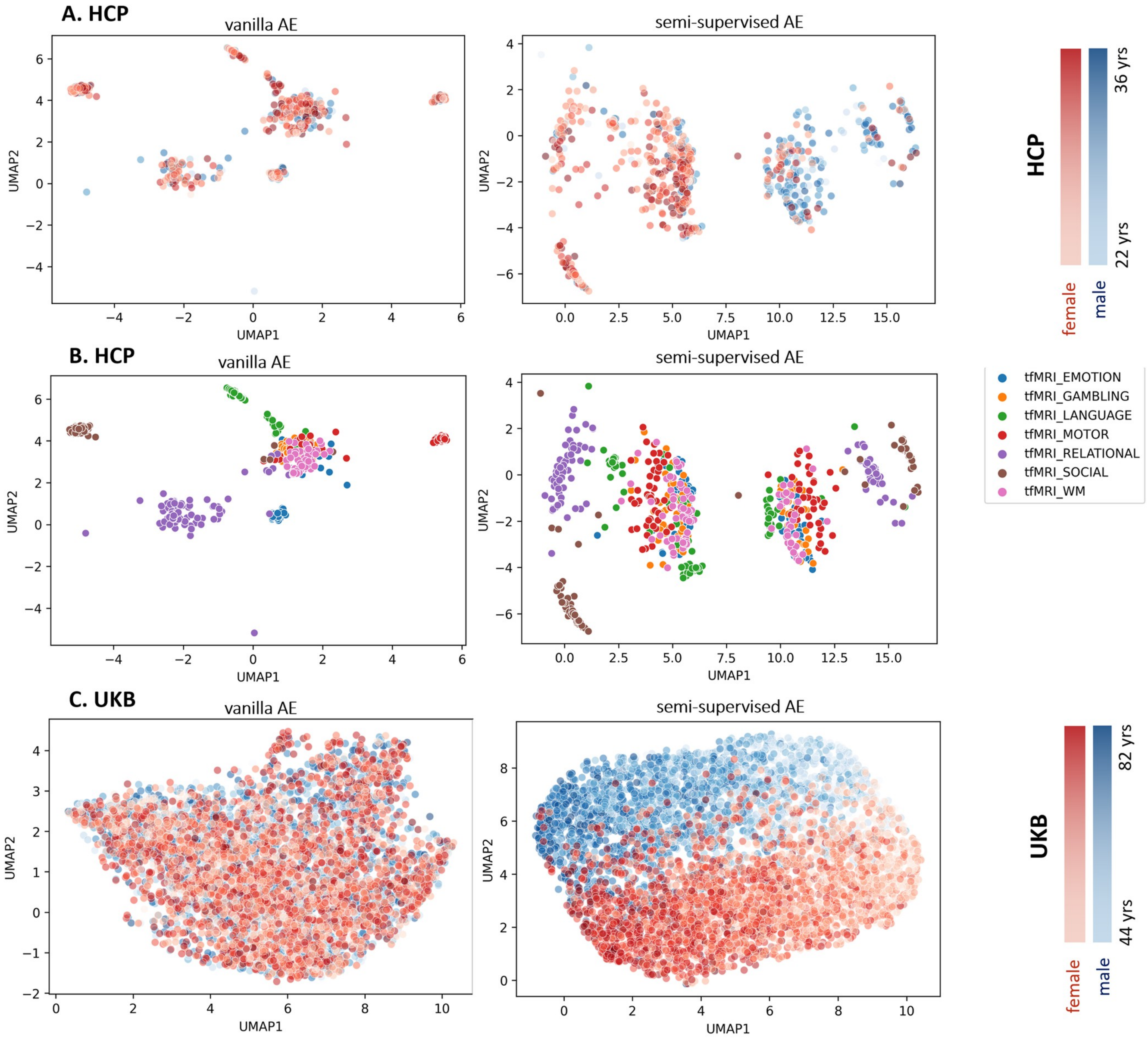

Fig 2. A) UMAP representation of the latent space of *selected contrasts from the Human Connectome Project (HCP) dataset according to Barch 2013*, *colored to show* age and sex separation. B) UMAP representation of the latent space of HCP task contrasts, showing task separation. This is identical to panel A, except that the data points are colored according to task instead of age and sex C) UMAP representation of the latent space of the Faces-Shapes task contrast in the UKB dataset.

https://doi.org/10.1371/journal.pone.0308329.g002

space, while still providing reasonable separation of the HCP tasks with respect to one another. The relationship with age is especially evident in UKBiobank, where the age range is wider.

Projection the latent representations to brain images

By navigating through the latent space, we are able to generate various distinct, yet meaningful, image representations and distill topological relationships that may not be evident in the original high dimensional space. This process is akin to moving along a manifold, where each point within this latent space correlates with a unique image and the proximity between points signifies the similarity (or conversely, distinctiveness) of the corresponding images. In this sense, the autoencoder acts as a translator, adeptly converting the complex, high-dimensional input data into a simpler, lower-dimensional, and more comprehensible format, while still preserving topological relationships evident in the high-dimensional data. Fig 3A shows how the activation changes in the input space by moving through the latent space for Motor Control task in HCP. This demonstrates that (i) as expected, the cue condition is the most distinctive from the other contrasts and exhibits high levels of inter-individual variability; (ii) the different motor task contrasts show more of a discrete and mostly more compact representation, with some evidence for higher functional differentiation in the left rather than the right hand. Since most of the HCP participants are right-handed, we interpret this as reflecting a more widespread neural activation pattern for the dominant, relative to the non-dominant hand.

We then map the spatial distribution of the different tasks learned by the autoencoder by projecting the cluster center of several HCP task contrasts (i.e., the main task contrasts reported in Barch et al 2013 [20]) and the Faces-Shapes contrast of UKB to the input space. We compared these synthetic activation patterns (Fig 3B and 3C) with the expected task labels derived from meta-analyses of existing findings using the NeuroSynth [74] meta-analytic database. Table 2 shows the tasks labels having the highest association with previous meta-analytic findings. (See S6 Fig in S1 File for the projection of the center of the UMAP of UKB latent space. While the mapping is not perfect, it is clear that many tasks strongly correspond to the expected cognitive domains (e.g., motor, working memory), although for other tasks the mapping is more ambiguous (e.g., social, gambling).

Association between latent variables and non-imaging covariates

Next, we aim to show the utility of the latent space in providing biomarkers on the basis of normative models. Fig 4 displays normative models applied to the UMAP representation of latent variable (see S7 Fig in S1 File for measures of fit for the HBR model). The distribution of the UMAP representations have a complex and non-Gaussian distribution (See qq-plots in S7 Fig in S1 File), which can be effectively modeled by leveraging the flexibility of HBR model. For example, in the left panels the distribution is negatively skewed for younger ages but becomes positively skewed for ages increase.

Fig 5 shows the Manhattan plot of the p-value of univariate correlation between non-imaging measures and the deviations from the normative model fit on the latent index of semi-supervised AE and PCA. This indicates strong associations with many nIDPs even after properly accounting for age and sex using the normative model. Furthermore, these associations are considerably stronger using semi-supervised autoencoder compared to the unsupervised representation provided by PCA. See S8 Fig in S1 File for the effect size of the associations between nIDPs and the latent index. We show the correlation of the raw UMAP scores with nIDPs in S9 Fig in S1 File (i.e. without first fitting a normative model to the latent variables). These associations are also strong but we consider this result with caution because it is potentially partially due to the high level of correlation between age, sex, and other cognitive and behavioral measures. Rather, we consider it preferable to interpret the deviations from a model that first removes the confounding effects of age and sex from the latent variables through the application of normative modeling.

Finally, we also performed comparisons of our approach with several alternative baseline methods, namely (i) a vanilla PCA on the image data, (ii) a two-stage PCA procedure that includes age and sex, thereby aiming to provide a linear benchmark for the semi-supervised autoencoder (iii) independent components analysis (ICA) and (iv) anatomically defined regions of interest. In summary, these approaches all resulted in considerably poorer

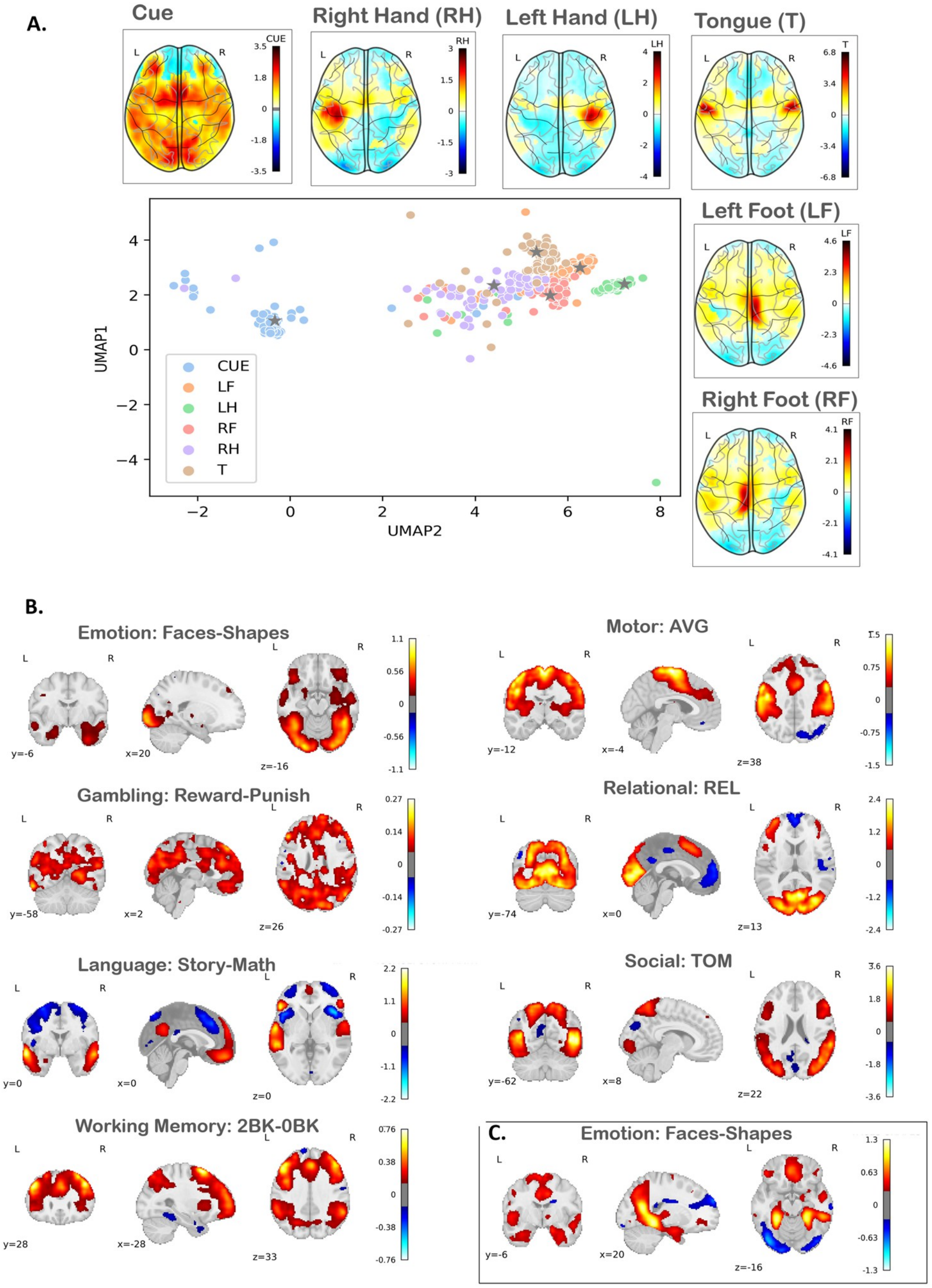

Fig 3. A) **Generated Motor-task contrast**: Changes in the input space activation correspond to moving through the centroid of the latent space for HCP Motor Control subtasks. B) **Generated HCP contrast**: Projections of the centers of the latent representation contrasts (according to Barch 2013) into the input image space. C) **Generated UKB contrast**: Projections of the centers of the latent representation of the Faces-Shapes subtask into the input image space. Abbreviations: LF: left foot; LH: left hand; RF: right foot; RH: right hand; T: tongue.

https://doi.org/10.1371/journal.pone.0308329.g003

Table 2. Correlation of activations at the center of latent in the original image space with previous findings, as derived from Neurosynth meta-analytic database. Each row represents a specific reported task in existing literature, along with the corresponding correlation values for each subtask in HCP and UKB datasets. Note that the corresponding subtasks per task are: Emotion (Faces-Shapes), Gambling (Reward-Punish), Language (Story-Math), Motor (AVG), Relational (REL), Social (Theory of Mind), and Working Memory (2BK-0BK).

Emotion (HCP)	**corr.**	**Gambling (HCP)**	**corr.**	**Language (HCP)**	**corr.**
visual	0.463	memory processing	0.180	social	0.367
face	0.462	decision making	0.176	theory mind	0.334
objects	0.387	visuo-spatial imaginary	0.174	mind	0.298
faces	0.385	mental state	0.135	listening	0.281
Motor (HCP)	**corr.**	**Relational (HCP)**	**Corr.**	**Social (HCP)**	**corr.**
motor	0.515	Visual	0.557	visual	0.517
movement	0.488	sensory perception	0.36	motion	0.458
touch	0.484	Task	0.336	object	0.384
voluntary movement	0.443	working memory	0.302	visuospatial cognition	0.374
Working Memory (HCP)	**corr.**	**Emotion (UKB)**	**corr.**		
working memory	0.384	spatial orientation	0.285		
working	0.38	recognition words	0.269		
task	0.371	Learning	0.176		
inhibitory control	0.37	visual information	0.159		

https://doi.org/10.1371/journal.pone.0308329.t002

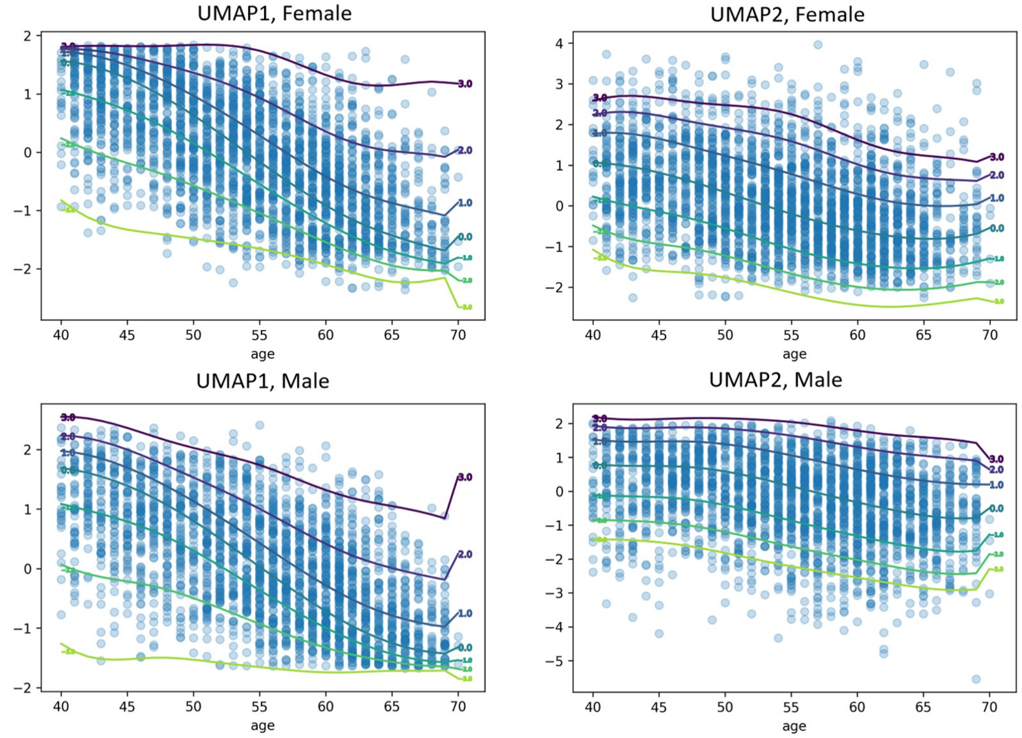

Fig 4. Normative models of the latent space UMAP components for males and females. The figure presents four separate normative models for two UMAP components of the latent space, each depicting the relationship between age and the corresponding UMAP component. The individualized deviations from the normative range represent the latent representation index. Each percentile line within each figure displays the level of deviation from the normative range for each time point, illustrating the degree to which individuals differ from the expected normative pattern across various percentiles.

https://doi.org/10.1371/journal.pone.0308329.g004

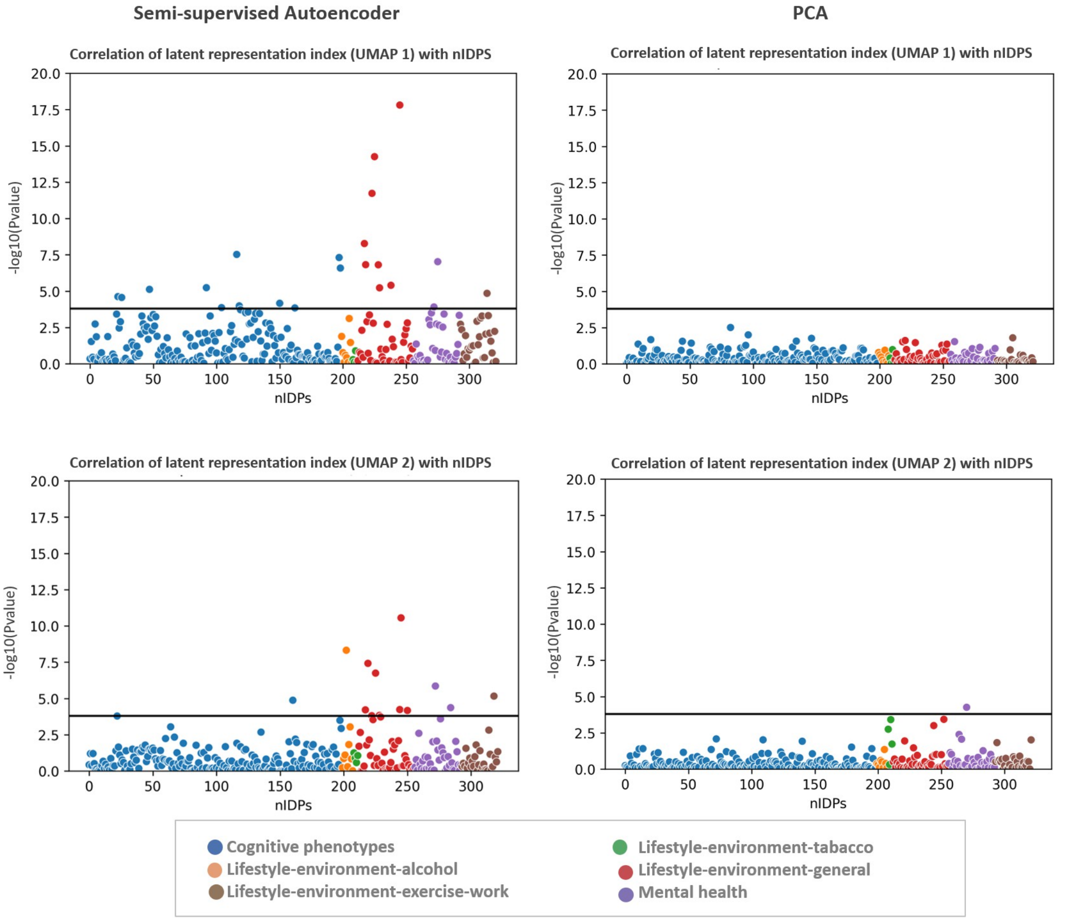

Fig 5. Manhattan plot of p-value of univariate correlation of non-imaging measures with the individualized deviations from normative UMAPs of latent space (latent representation index) and PCA. The black line is Bonferroni-corrected p-value threshold.

https://doi.org/10.1371/journal.pone.0308329.g005

performance and weak associations with non-imaging measures (see S4 Table and S10-S13 Figs in S1 File).

Discussion

In this study, our primary objective was to develop a 3-D convolutional autoencoder architecture capable of extracting more informative, lower-dimensional latent representations of brain activity from task-based fMRI data. These representations better capture the underlying neural processes associated with specific cognitive tasks and demonstrate the potential to predict behavioral and cognitive measures. While the concept of lower-dimensional representations and their relation to behavior is not novel, our work contributes by tailoring the autoencoder-based model to fMRI data, providing interpretable latent representations that can be compared to linear methods like PCA. Importantly, through our joint optimization framework, we are able to control the representations that the autoencoder learns, in this case, learning a manifold related to age and sex, where the deviations from this manifold are informative about individual differences and provide candidate biomarkers. By "interpretable" and "meaningful latent representations," we refer to the ability of these representations to be effectively mapped to non-imaging variables and to be understood both within the latent space and in the original voxel space. This dual interpretability ensures that the latent space not only captures essential features relevant to the imaging data but also maintains a clear relationship with biological and clinical variables. As a result, decoding a latent representation yields outputs that are coherent with the original data, ensuring that the generated outputs are biologically plausible and clinically relevant. We argue that this capability is essential for validating the model's predictions and ensuring that the latent representations are effective for further analyses, such as predicting cognitive or behavioral outcomes.

Our model allows for accurate reconstruction of the data while representing demographic variation and presents methods to visualize, interpret, and control the learned latent space representation. We show a simple example of how this approach can be used to understand topological relationships of different task representations in the brain using the HCP motor task. By defining a latent index, we established the utility of this approach for developing biomarkers that predict behavioral measures. We demonstrated that our model learns salient features that capture age, other sources of population stratification, and are strongly associated with clinical and behavioral features.

Learning a generic latent representation

While our approach can be applied to various types of imaging data, we chose to focus on task-based fMRI contrasts in this study. Task-based fMRI is easily relatable to specific cognitive functions and has a long history in the neuroimaging field as suitable method for studying associations between brain and behavior. However, our approach could also be adopted for other types of neuroimaging-derived readouts, although this might require a different architecture than the one we used here. More specifically, the HCP task-fMRI data enabled us to estimate a generic latent space representation across diverse cognitive tasks [19, 20], whilst also providing good whole brain coverage across all the tasks [20]. During the training, this mapping allows the autoencoder to learn various activation patterns across the brain, rather than focusing on specific task-related effects that might be localized to particular brain regions.

Mapping the latent space

Considering the limited age range and the relatively small number of test participants in each HCP model (N≈95), the effect of age and sex in the latent space is not clear, although the relationships between tasks (Fig 2B) and different conditions within the same task (Fig 2A) are clearly distinguishable. In contrast, the UMAP of UKB generates a clear age continuum and good separation in terms of sex. This indicates that moving from one point in the manifold to another can be meaningfully traced back through the input space and that changes within the latent space reflect salient changes within the input data, although the precise relationships may be different across datasets, depending on the degree to which different modeled variables are reflected in the original data. This is in line with the standard interpretation of autoencoders as learning a manifold [49].

Unsupervised training of the model yields interpretable representations where different task contrasts are well-clustered and separated in the latent space. In contrast, with the semi-supervised learning, our representation is tailored to focus on mapping the latent space back to the original space, with the added control of age and sex. This approach ensures that the latent manifold clearly reflects individual differences related to the demographic variables in the underlying imaging data, whilst still providing good separation in terms of task contrasts.

Projection of latent representation to original space: For the majority of the contrasts and particularly language (story-math), working memory (2BK-0BK), and motor control (AVG), the projection of the center of the latent space to the original scan image space faithfully represents the group maps in [20]. In the context of interpretability of findings, the meaningful projection of the latent space can be viewed as an example of explainable AI in complex models.

Association of the latent representation index with nIDPs measures

A key aspect of summarizing the complex spatial maps of tfMRI is to preserve individual variability. To complement this, these summaries or representations should contain biological information that can be linked to cognitive, behavioral and clinical characteristics. Since the latent space also represents age and sex, and because age is strongly associated with a variety of cognitive and behavioral scores, the correlation of latent variables and nIDPs may be disrupted by the confounding effect of age (see S8 Fig in S1 File for the correlation of UMAPs and nIPDs). To disentangle clinically relevant variation from variation due to age and sex from the UMAP representation, we applied normative modeling based on hierarchical Bayesian regression. Here, the individualized deviations or latent representation index indicates the distance from the normative latent variables transformed by UMAP. We showed that this index is strongly associated with several nIDP scores after accounting for confounding variables (age and sex). Hence, the notion of normative latent variables may provide the basis for the development of a biomarker that predicts cognitive and behavioral characteristics. Moreover, we show that this association is stronger than classical linear models such as PCA.

Network architecture

The architectural hyper-parameters of the autoencoder were chosen during the pilot study, solely based on model's performance in terms of the reconstruction error; no other readouts i.e., non-imaging measures, were used for evaluations. Additionally, the data used for the pilot study were not reused. Some decisions about the network structure were made before estimating the model. For example, to preserve the morphology of the images and thus improve the interpretability, we decided to use a 3-D convolutional network [30–32, 45]. In order to control over of latent space, we used a dense layer in the bottleneck of the autoencoder [49].

We designed our autoencoder with the specific nature of our high-dimensional neuroimaging data in mind, imposing several constraints on the model beforehand. For example, the networks evaluated were not particularly deep. To reduce the memory usage and computational complexity, we took advantage of the weight sharing of convolutional layers. We aimed to identify low-level features that may be translation invariant, with a significant benefit being the ability to scale the networks to whole-brain data [75]. The kernel size was set to be 3×3×3 to retain the details of the downsampled image scans. Average pooling layers were positioned after each convolutional layer to smooth sharp features, reduce the number of parameters, and minimize the chance of overfitting. We relied on the pilot study to select the remaining model parameters, such as the number of filters.

We assigned both unsupervised (image reconstruction error) and supervised (age and sex prediction) loss functions to our semi-supervised AE, aiming to find meaningful latent representations of data that can be mapped to the non-imaging variables and interpreted both in the latent space and in the original voxel space. Our model showed high performance in predicting age and sex. The contribution of supervised and unsupervised loss can be also redefined in order to emphasize the optimization process. This results in a semi-supervised setting that allows the latent space to partially encode specific features of the data [8]. Another interesting future direction is to train an autoencoder to predict different data (e.g., a follow-up timepoint in longitudinal studies). This would sensitize the latent space to changes relevant to aging or pathology, suggesting that the latent representation may also be useful for generating features for downstream analyses aiming at predicting these changes.

Limitations and future work

The increased number of neuroimaging scans provides a unique opportunity to transcend linear mappings, but it is also necessary to acknowledge some limitations. Traditional image processing techniques often used in deep learning are not entirely applicable here. For example, while data augmentation methods such as image mirroring, flipping, skewing, or segmenting are straightforward approaches to increase the number of samples and have been applied in neuroimaging applications [11], we did not consider them appropriate here. Such augmentation strategies do not faithfully preserve invariances known to occur in the brain, such as the lateralization of brain functions (e.g. the association of left lateralization in language processing [76]).

The generalizability of the latent representation is another important concern that has not been fully addressed here due to age range differences between two datasets. UKB contains the Hariri faces-shapes emotion task [77], which is very similar to the emotion task of HCP (effectively a shorter version). Although the common contrasts provide a great opportunity for further validation of the model, the age gap limits the capacity to test the generalization of the latent space across cohorts.

Computational complexity is another limitation. Training an autoencoder on large neuroimaging data is computationally more demanding compared to linear models. In this work, we set the trade-off parameter (lambda) governing the contribution of supervised and unsupervised loss components in a relatively informal manner. A quantitative evaluation would have required us to define the relative value of each component (e.g. how much to favor prediction of the supervised targets over reconstruction or vice versa). It is possible that more careful optimization of this parameter may yield improved performance.

In our study, we focused on age and sex as supervised factors in our model, as they are well-known demographic variables with significant effects on brain structure and function. By incorporating these factors, we aimed to disentangle the age- and sex-related variations from other sources of variability in the data. However, we acknowledge that there may be other factors contributing to the differences between the HCP and UK Biobank samples, such as socioeconomic status, education, and genetic predispositions. These factors could also influence the latent representations and their generalizability. In future work, it would be valuable to extend our model to incorporate additional factors and investigate how they contribute to the observed inter-individual differences in brain activity. This extension will help provide a more comprehensive understanding of the factors influencing the generalizability of the latent representations and their relationships with brain-behavior associations.

Furthermore, we recognize the importance of comparing our approach to alternative methods and demonstrating its potential applicability in clinical practice. Our study compares our model to PCA, ICA, and ROIs , showing that the autoencoder-based approach better predicts behavioral and cognitive measures. We also believe our model holds promise for clinical practice, as it may help identify atypical patterns of brain activity not easily detected by traditional behavioral measures. This could be particularly relevant for large-scale datasets like the UK Biobank, where neuroimaging data can be leveraged to inform personalized interventions and improve clinical outcomes, ultimately advancing our understanding of the links between the brain and behavior.

In this work we principally used the very low-level representation derived from the UMAP, however for some analyses it may be preferable to use the ultimately richer representation derived from the autoencoder itself. Such an approach may be particularly valuable to relate the learned representation to external variables using multivariate methods such as classifiers or canonical correlation analysis.

Conclusion

Here, we applied 3-dimensional autoencoder to two large-scale datasets to find an interpretable latent representation of high dimensional task fMRI image data by controlling for demographic information. We applied normative modeling to the latent variables to define an index to find a mapping to non-imaging measures.

Our model showed high performance in terms of age and sex prediction and was capable of capturing complex biological, cognitive, and clinical characteristics while preserving the individualized variabilities using a latent representation index. We consider this representation to provide an excellent basis for understanding inter-individual differences in neural representations in clinical and fundamental neuroscience research.

Supporting information

S1 File.
(PDF)

Author Contributions

Conceptualization: Mariam Zabihi, Seyed Mostafa Kia, Thomas Wolfers, Stijn de Boer, Charlotte Fraza, Richard Dinga, Alberto Llera Arenas, Danilo Bzdok, Christian F. Beckmann, Andre Marquand.

Formal analysis: Mariam Zabihi.

Methodology: Mariam Zabihi, Seyed Mostafa Kia, Andre Marquand.

Supervision: Christian F. Beckmann, Andre Marquand.

Visualization: Mariam Zabihi.

Writing – original draft: Mariam Zabihi.

Writing – review & editing: Andre Marquand.

References

1. Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, et al. Scanning the horizon: Towards transparent and reproducible neuroimaging research. Nat Rev Neurosci. 2017; 18 (2):115–26. https://doi.org/10.1038/nrn.2016.167 PMID: 28053326
2. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017; 42(December 2012):60–88. https://doi.org/10.1016/j.media.2017.07.005 PMID: 28778026
3. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in Neuroimaging [Internet]. Vol. 12, Neuroinformatics. Humana Press Inc.; 2014 [cited 2021 Jan 6]. p. 229–44. Available from: /pmc/articles/PMC4040248/?report=abstract https://doi.org/10.1007/s12021-013-9204-3 PMID: 24013948
4. Korolev S, Safiullin A, Belyaev M, Dodonova Y. Residual and plain convolutional neural networks for 3D brain MRI classification. Proc - Int Symp Biomed Imaging. 2017;835–8.
5. Gong W, Beckmann CF, Smith SM. Phenotype Discovery from Population Brain Imaging. bioRxiv [Internet]. 2020; Available from: https://www.biorxiv.org/content/early/2020/03/05/2020.03.05.973172
6. Bellman RE. Adaptive control processes: a guided tour. Vol. 2045. Princeton university press; 2015.
7. Schulz MA, Yeo BTT, Vogelstein JT, Mourao-Miranada J, Kather JN, Kording K, et al. Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. Nat Commun [Internet]. 2020; 11(1). Available from: https://doi.org/10.1038/s41467-020-18037-z PMID: 32843633

8. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. Hum Brain Mapp [Internet]. 2019 Feb 15 [cited 2019 Jun 6]; 40(3):944–54. Available from: http://doi.wiley.com/10.1002/hbm.24423 https://doi.org/10.1002/hbm.24423 PMID: 30311316

9. Pinaya WHL, Scarpazza C, Garcia-Dias R, Vieira S, Baecker L, da Costa PF, et al. Normative modelling using deep autoencoders: a multi-cohort study on mild cognitive impairment and Alzheimer's disease. bioRxiv. 2020;

10. Cole JH, Poudel RPK, Tsagkrasoulis D, Caan MWA, Steves C, Spector TD, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage. 2017; 163:115–24. https://doi.org/10.1016/j.neuroimage.2017.07.059 PMID: 28765056

11. Peng H, Gong W, Beckmann CF, Vedaldi A, Smith SM. Accurate brain age prediction with lightweight deep neural networks. Med Image Anal. 2021 Feb 1; 68:101871. https://doi.org/10.1016/j.media.2020.101871 PMID: 33197716

12. Dinsdale NK, Bluemke E, Smith SM, Arya Z, Vidaurre D, Jenkinson M, et al. Learning patterns of the ageing brain in MRI using deep convolutional networks. Neuroimage. 2021 Jan 1; 224:117401. https://doi.org/10.1016/j.neuroimage.2020.117401 PMID: 32979523

13. Kiesow H, Spreng RN, Holmes AJ, Chakravarty MM, Marquand AF, Yeo BTT, et al. Hidden population modes in social brain morphology: Its parts are more than its sum. bioRxiv [Internet]. 2020 Aug 7 [cited 2021 Jan 6];2020.08.07.241497. Available from: https://doi.org/10.1101/2020.08.07.241497

14. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2015.

15. Pinaya WHL, Scarpazza C, Garcia-Dias R, Vieira S, Baecker L, F da Costa P, et al. Using normative modelling to detect disease progression in mild cognitive impairment and Alzheimer's disease in a cross-sectional multi-cohort study. Sci Rep [Internet]. 2021 Dec 1 [cited 2022 Feb 1]; 11(1). Available from: https://pubmed.ncbi.nlm.nih.gov/34344910/

16. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neurosci Biobehav Rev [Internet]. 2017 Mar 1 [cited 2019 Jun 6]; 74:58–75. Available from: https://www.sciencedirect.com/science/article/pii/S0149763416305176?via%3Dihub https://doi.org/10.1016/j.neubiorev.2017.01.002 PMID: 28087243

17. Hao AJ, He BL, Yin CH. Discrimination of ADHD children based on Deep Bayesian Network. IET Conf Publ. 2015; 2015(CP680).

18. Wolfers T, Doan N, Kaufmann T, al et. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. JAMA Psychiatry [Internet]. 2018 Oct 10; Available from: https://doi.org/10.1001/jamapsychiatry.2018.2467

19. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: An overview. Neuroimage. 2013; 80:62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041 PMID: 23684880

20. Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, et al. Function in the human connectome: Task-fMRI and individual differences in behavior. Neuroimage. 2013; 80:169–89. https://doi.org/10.1016/j.neuroimage.2013.05.033 PMID: 23684877

21. Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior [Internet]. Vol. 18, Nature Neuroscience. Nature Publishing Group; 2015 [cited 2020 Sep 1]. p. 1565–7. Available from: http://www.nature.com/ https://doi.org/10.1038/nn.4125 PMID: 26414616

22. Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci [Internet]. 2015; 18 (October):1–11. Available from: https://doi.org/10.1038/nn.4135 PMID: 26457551

23. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci. 2016; 19(11):1523–36. https://doi.org/10.1038/nn.4393 PMID: 27643430

24. Gupta L, Besseling RMH, Overvliet GM, Hofman PAM, De Louw A, Vaessen MJ, et al. Spatial heterogeneity analysis of brain activation in fMRI. NeuroImage Clin [Internet]. 2014; 5:266–76. Available from: https://doi.org/10.1016/j.nicl.2014.06.013 PMID: 25161893

25. Burgess GC, Gray JR, Conway ARA, Braver TS. Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. J Exp Psychol Gen. 2011; https://doi.org/10.1037/a0024695 PMID: 21787103

26. Bzdok D. Classical statistics and statistical learning in imaging neuroscience [Internet]. Vol. 11, Frontiers in Neuroscience. Frontiers Media S.A.; 2017 [cited 2021 Feb 17]. p. 543. Available from: www.frontiersin.org https://doi.org/10.3389/fnins.2017.00543 PMID: 29056896

27. Suk H Il, Lee SW, Shen D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct Funct [Internet]. 2015 [cited 2021 Jan 6]; 220(2):841–59. Available from: /pmc/articles/PMC4065852/?report=abstract https://doi.org/10.1007/s00429-013-0687-3 PMID: 24363140

28. Davatzikos C. Machine learning in neuroimaging: Progress and challenges. Neuroimage. 2019; 197:652. https://doi.org/10.1016/j.neuroimage.2018.10.003 PMID: 30296563

29. He T, Kong R, Holmes AJ, Nguyen M, Sabuncu MR, Eickhoff SB, et al. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. Neuroimage [Internet]. 2020 Feb 1 [cited 2021 Jan 29];206. Available from: https://pubmed.ncbi.nlm.nih.gov/31610298/ https://doi.org/10.1016/j.neuroimage.2019.116276 PMID: 31610298

30. Tudosiu PD, Varsavsky T, Shaw R, Graham M, Nachev P, Ourselin S, et al. Neuromorphologicaly-preserving Volumetric data encoding using VQ-VAE. arXiv [Internet]. 2020 Feb 13 [cited 2021 Jan 6];1–13. Available from: http://arxiv.org/abs/2002.05692

31. Kwon G, Han C, Kim D shik. Generation of 3D Brain MRI Using Auto-Encoding Generative Adversarial Networks. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) [Internet]. 2019 Aug 7 [cited 2021 Jan 6];11766 LNCS:118–26. Available from: http://arxiv.org/abs/1908.02498

32. Choi H, Kang H, Lee DS. Predicting aging of brain metabolic topography using variational autoencoder. Front Aging Neurosci [Internet]. 2018 Jul 12 [cited 2021 Jan 6]; 10(JUL):212. Available from: /pmc/articles/PMC6052253/?report=abstract https://doi.org/10.3389/fnagi.2018.00212 PMID: 30050430

33. Huang H, Hu X, Zhao Y, Makkie M, Dong Q, Zhao S, et al. Modeling Task fMRI Data Via Deep Convolutional Autoencoder. IEEE Trans Med Imaging. 2018 Jul 1; 37(7):1551–61. https://doi.org/10.1109/TMI.2017.2715285 PMID: 28641247

34. Brown JA, Lee AJ, Pasquini L, Seeley WW. A dynamic gradient architecture generates brain activity states. Neuroimage. 2022; 261. https://doi.org/10.1016/j.neuroimage.2022.119526 PMID: 35914669

35. Kim JH, Zhang Y, Han K, Wen Z, Choi M, Liu Z. Representation learning of resting state fMRI with variational autoencoder. Neuroimage. 2021; 241. https://doi.org/10.1016/j.neuroimage.2021.118423 PMID: 34303794

36. Cui Y, Zhao S, Chen Y, Han J, Guo L, Xie L, et al. Modeling brain diverse and complex hemodynamic response patterns via deep recurrent autoencoder. IEEE Trans Cogn Dev Syst. 2020; 12(4).

37. Kim JH, De Asis-Cruz J, Krishnamurthy D, Limperopoulos C. Toward a more informative representation of the fetal–neonatal brain connectome using variational autoencoder. Elife. 2023; 12.

38. Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. Neuroimage [Internet]. 2011 Apr 15 [cited 2021 Jan 6]; 55 (4):1519–27. Available from: https://pubmed.ncbi.nlm.nih.gov/21167288/ https://doi.org/10.1016/j.neuroimage.2010.12.028 PMID: 21167288

39. Sidhu G, Asgarian N, Greiner R, Brown MRG. Kernel principal component analysis for dimensionality reduction in fMRI-based diagnosis of ADHD. Front Syst Neurosci [Internet]. 2012 Oct 10 [cited 2021 Jan 6]; 6(October):1–17. Available from: https://pubmed.ncbi.nlm.nih.gov/23162439/ https://doi.org/10.3389/fnsys.2012.00074 PMID: 23162439

40. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. Neuroimage. 2017 Jan 15; 145:166–79. https://doi.org/10.1016/j.neuroimage.2016.10.038 PMID: 27989847

41. Calhoun VD, Liu J, Adali T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage [Internet]. 2009 [cited 2021 Jan 6];45(1 Suppl). Available from: https://pubmed.ncbi.nlm.nih.gov/19059344/ https://doi.org/10.1016/j.neuroimage.2008.10.057 PMID: 19059344

42. Thirion B, Faugeras O. Dynamical components analysis of fMRI data through kernel PCA. Neuroimage. 2003 Sep 1; 20(1):34–49. https://doi.org/10.1016/s1053-8119(03)00316-1 PMID: 14527568

43. Bzdok D, Yeo BTT. Inference in the age of big data: Future perspectives on neuroscience [Internet]. Vol. 155, NeuroImage. Academic Press Inc.; 2017 [cited 2021 Feb 17]. p. 549–64. Available from: https://pubmed.ncbi.nlm.nih.gov/28456584/ https://doi.org/10.1016/j.neuroimage.2017.04.061 PMID: 28456584

44. Smith SM, Nichols TE. Statistical Challenges in "Big Data" Human Neuroimaging [Internet]. Vol. 97, Neuron. Cell Press; 2018 [cited 2021 Feb 18]. p. 263–8. Available from: https://pubmed.ncbi.nlm.nih.gov/29346749/ https://doi.org/10.1016/j.neuron.2017.12.018 PMID: 29346749

45. Payan A, Montana G. Predicting Alzheimer ' s disease: a neuroimaging study with 3D convolutional neural networks. arXiv Prepr arXiv150202506. 2015;1–9.

46. Suk H Il, Lee SW, Shen D. Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal. 2017 Apr 1; 37:101–13. https://doi.org/10.1016/j.media.2017.01.008 PMID: 28167394

47. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018 Jul 1; 50(7):912–9. https://doi.org/10.1038/s41588-018-0152-6 PMID: 29942086

48. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) [Internet]. Springer Verlag; 2019 [cited 2021 Jan 6]. p. 311–20. Available from: https://doi.org/10.1007/978-3-030-11726-9_28

49. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Vol. 1. MIT press Cambridge; 2016.

50. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. Vol. 24, Molecular Psychiatry. Nature Publishing Group; 2019. p. 1415–24. https://doi.org/10.1038/s41380-019-0441-1 PMID: 31201374

51. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. Biol Psychiatry Cogn Neurosci Neuroimaging [Internet]. 2016; 1(5):433–47. Available from: https://doi.org/10.1016/j.bpsc.2016.04.002 PMID: 27642641

52. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. Biol Psychiatry. 2016; 80(7):552–61. https://doi.org/10.1016/j.biopsych.2015.12.023 PMID: 26927419

53. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun [Internet]. 2020; 11(1):2624. Available from: https://doi.org/10.1038/s41467-020-15948-9

54. Bzdok D, Varoquaux G, Grisel O, Eickenberg M, Poupon C, Thirion B. Formal Models of the Network Co-occurrence Underlying Mental Operations. Bassett DS, editor. PLOS Comput Biol [Internet]. 2016 Jun 16 [cited 2020 Aug 11]; 12(6):e1004994. Available from: https://dx.plos.org/10.1371/journal.pcbi.1004994 https://doi.org/10.1371/journal.pcbi.1004994 PMID: 27310288

55. Bzdok D, Eickenberg M, Grisel O, Thirion B, Varoquaux Semi G, Varoquaux G. Semi-Supervised Factored Logistic Regression for High-Dimensional Neuroimaging Data [Internet]. 2015 [cited 2021 Feb 17]. Available from: https://hal.archives-ouvertes.fr/hal-01211248

56. Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat Neurosci [Internet]. 2015; 18(11):1565–7. Available from: https://doi.org/10.1038/nn.4125 PMID: 26414616

57. Marquand AF, Haak K V., Beckmann CF. Functional corticostriatal connection topographies predict goal-directed behaviour in humans. Nat Hum Behav [Internet]. 2017 Jul 24 [cited 2021 Feb 10]; 1 (8):146. Available from: www.nature.com/nhumbehav https://doi.org/10.1038/s41562-017-0146 PMID: 28804783

58. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage [Internet]. 2013 Oct 5 [cited 2021 Jan 22]; 80:105–24. Available from: https://pubmed.ncbi.nlm.nih.gov/23668970/ https://doi.org/10.1016/j.neuroimage.2013.04.127 PMID: 23668970

59. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage [Internet]. 2018; 166:400–24. Available from: https://www.sciencedirect.com/science/article/pii/S1053811917308613 https://doi.org/10.1016/j.neuroimage.2017.10.034 PMID: 29079522

60. Fred Agarap AM. Deep Learning using Rectified Linear Units (ReLU). [cited 2022 Jun 11]; Available from: https://github.com/AFAgarap/relu-classifier.

61. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014; 15.

62. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. 2008.

63. Leonardsen EH, Peng H, Kaufmann T, Agartz I, Andreassen OA, Celius EG, et al. Deep neural networks learn general and clinically relevant representations of the ageing brain. Neuroimage. 2022 Aug 1; 256:119210. https://doi.org/10.1016/j.neuroimage.2022.119210 PMID: 35462035

64. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Journal of Machine Learning Research. 2010.

65. Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. 2015.

66. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [Internet]. 2018 Feb 9 [cited 2021 Jan 6]; Available from: http://arxiv.org/abs/1802.03426

67. der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008; 9(11).

68. Gordon EM, Chauvin RJ, Van AN, Rajesh A, Nielsen A, Newbold DJ, et al. A somato-cognitive action network alternates with effector regions in motor cortex. Nat 2023 6177960 [Internet]. 2023 Apr 19 [cited 2023 Jul 3]; 617(7960):351–9. Available from: https://www.nature.com/articles/s41586-023-05964-2 https://doi.org/10.1038/s41586-023-05964-2 PMID: 37076628

69. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. Biol Psychiatry Cogn Neurosci Neuroimaging [Internet]. 2019 Jun; 4(6):567–78. Available from: https://linkinghub.elsevier.com/retrieve/pii/S245190221830329X

70. Rutherford S, Fraza C, Dinga R, Mostafa Kia S, Wolfers T, Zabihi M, et al. Charting brain growth and aging at high spatial precision. 2022; 11:72904.

71. Kia SM, Huijsdens H, Dinga R, Wolfers T, Mennes M, Andreassen OA, et al. Hierarchical Bayesian Regression for Multi-site Normative Modeling of Neuroimaging Data. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) [Internet]. Springer Science and Business Media Deutschland GmbH; 2020 [cited 2021 May 10]. p. 699–709. Available from: https://doi.org/10.1007/978-3-030-59728-3_68

72. Fraza CJ, Dinga R, Beckmann CF, Marquand AF. Warped Bayesian linear regression for normative modelling of big data. Neuroimage. 2021 Dec 15; 245:118715. https://doi.org/10.1016/j.neuroimage.2021.118715 PMID: 34798518

73. Boer AAA de, Kia SM, Rutherford S, Zabihi M, Fraza C, Barkema P, et al. Non-Gaussian Normative Modelling With Hierarchical Bayesian Regression. bioRxiv. 2022;

74. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. Nat Methods. 2011; 8(8). https://doi.org/10.1038/nmeth.1635 PMID: 21706013

75. Lecun Y, Bengio Y, Hinton G. Deep learning [Internet]. Vol. 521, Nature. Nature Publishing Group; 2015 [cited 2021 Jan 6]. p. 436–44. Available from: http://colah.github.io/ https://doi.org/10.1038/nature14539 PMID: 26017442

76. Frost JA, Binder JR, Springer JA, Hammeke TA, Bellgowan PSF, Rao SM, et al. Language processing is strongly left lateralized in both sexes. Evidence from functional MRI. Brain [Internet]. 1999 [cited 2021 Jan 29]; 122(2):199–208. Available from: https://pubmed.ncbi.nlm.nih.gov/10071049/ https://doi.org/10.1093/brain/122.2.199 PMID: 10071049

77. Hariri AR, Tessitore A, Mattay VS, Fera F, Weinberger DR. The amygdala response to emotional stimuli: A comparison of faces and scenes. Neuroimage [Internet]. 2002 [cited 2021 Feb 11]; 17(1):317–23. Available from: https://pubmed.ncbi.nlm.nih.gov/12482086/ https://doi.org/10.1006/nimg.2002.1179 PMID: 12482086