Structure-aware neural networks in the study of multi-modal population cohorts : an application to mental health

Corentin Ambroise

université PARIS-SACLAY

Structure-aware neural networks in the study of multi-modal population cohorts: an application to mental health

Réseaux de neurones sensibles à la structure pour l'analyse de données multimodales de population : une application à la santé mentale

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 575 : electrical, optical, bio : physics and engineering (EOBE)
Spécialité de doctorat : Sciences de l'information et de la communication
Graduate School : Sciences de l'ingénierie et des systèmes. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **BAOBAB** (Université Paris-Saclay, CEA, CNRS), sous la direction de **Vincent FROUIN**, Directeur de recherche, le co-encadrement de **Antoine GRIGIS**, Docteur, Ingénieur de recherche

Thèse soutenue à Paris-Saclay, le 12 juin 2024, par

Corentin AMBROISE

Composition du jury

Membres du jury avec voix délibérative

Stéphanie ALLASSONNIERE Professeure des universités, Université Paris-Cité	Présidente
Diana MATEUS Professeure des universités, Ecole Centrale Nantes	Rapportrice & examinatrice
Janaina MOURAO-MIRANDA Professor (équiv. HDR), University College London	Rapportrice & examinatrice
Charles LAIDI Professeur des universités - practicien hospitalier, Université Paris-Est Créteil	Examinateur

THESE DE DOCTORAT
NNT : 2024UPAST065

université PARIS-SACLAY
ÉCOLE DOCTORALE
Physique et ingénierie : Electrons, Photons, Sciences du vivant (EOBE)

Titre: Réseaux de neurones sensibles à la structure pour l'analyse de données multimodales de population : une application à la santé mentale.

Mots clés: Réseaux de neurones, Multi-modal, Etudes de population, Graphes, Neuro-imagerie, Interpretabilité

Résumé: Il est actuellement reconnu que le fait de s'appuyer uniquement sur des stratégies de classification conventionnelles à partir d'une seule source de données n'est pas efficace pour comprendre, diagnostiquer ou pronostiquer les syndromes psychiatriques. Les objectifs de classification reposent simplement sur les étiquettes des cliniciens qui, à elles seules, n'expriment pas une très grande variabilité. En 2009, le Research Domain Criteria (RDoC) a recommandé une approche plus complète pour étudier les troubles psychiatriques en incorporant divers types de données qui couvrent différents niveaux d'organisation de la vie (par exemple, l'imagerie, la génétique, les symptômes). La proposition suggère qu'une description complète d'une pathologie nécessite la prise en compte de dimensions multiples, qui peuvent être partagées entre différents syndromes psychiatriques et même contribuer à la variabilité non pathologique. Des cadres efficaces pour l'apprentissage non supervisé, spécifiquement conçus pour des approches multivariées et multimodales, devraient offrir des méthodologies pour traiter et intégrer le type d'ensembles de données préconisé par le RDoC. L'apprentissage profond nous permet d'apprendre dans des environnements multimodaux avec une structure spécifique à la modalité et une structure de corrélation intermodale. Pour modéliser la structure intra-modalité, nous utilisons des réseaux de neurones convolutionels spécifiques, qui permettent d'apprendre à partir de mesures cérébrales corticales réparties sur un maillage sphérique et ainsi de révéler des biomarqueurs originaux. Dans ce contexte, nous proposons 5 augmentations de données et les appliquons dans l'un des nombreux nouveaux schémas d'apprentissage auto-supervisé reposant principalement sur l'augmentation de données. Ce travail permet à l'apprentissage par représentation profonde d'initialiser correctement le réseau sur d'énormes cohortes de patients sains, puis de le transférer pour étudier la pathologie clinique d'intérêt dans des cohortes plus petites. D'autre part, nous avons identifié les auto-encodeurs variationnels multi-vues comme de bons candidats pour intégrer des modalités multiples. En outre, nous remettons en question l'hypothèse courante selon laquelle les réseaux neuronaux ne sont pas interprétables. Nous utilisons une procédure d'avatar numérique comme module d'interprétabilité capable de rendre compte des relations inter-vues apprises au sein d'un auto-encodeur multi-vues. En particulier, nous intégrons cette procédure dans une nouvelle méthode qui combine plusieurs de ces modèles et interprétations, encapsulée dans une procédure de sélection par stabilité pour identifier des associations significatives et reproductibles entre les modalités d'imagerie cérébrale et le comportement. Nous appliquons cette méthode pour mettre en évidence des associations cerveau-comportement spécifiques présentes dans la cohorte transdiagnostique Healthy Brain Network (HBN). Les associations cerveau-comportement identifiées établissent des connexions entre les caractéristiques corticales régionales issues de l'imagerie par résonance magnétique structurelle et les dossiers cliniques électroniques évaluant les symptômes psychiatriques. Nous montrons que cette méthode est capable de trouver des associations pertinentes et stables.

Title: Structure-aware neural networks in the study of multi-modal population cohorts : an application to mental health.

Keywords: Neural networks, Multi-modal, Population study, Graphs, Neuro-imaging, Interpretability

Abstract: It is currently acknowledged that relying solely on conventional classification strategies from a single data source is not effective to understand, diagnose or prognose psychiatric syndromes. The classification targets simply rely on clinician labels that alone do not express a very large variability. In 2009, the Research Domain Criteria (RDoC) recommended a more comprehensive approach to study psychiatric disorders by incorporating diverse data types that cover various levels of life organization (e.g., imaging, genetic, symptoms). The proposal suggests that a thorough description of a pathology requires consideration of multiple dimensions, which may be shared across different psychiatric syndromes and even contribute to non-pathological variability. Efficient frameworks for unsupervised learning, specifically designed for multivariate and multimodal approaches, are anticipated to offer methodologies for handling and integrating the kind of datasets advocated by the RDoC. Deep learning allows us to learn in multimodal settings with modality-specific structure and intermodality correlation structure. To model intra-modality structure, we use specific convolutional neural networks that enable to learn from cortical brain measures distributed across a spherical mesh and thus reveal original biomarkers. In this context, we propose 5 data augmentations and apply them in one of the many novel self-supervised learning schemes relying moslty on data augmentation. This work allows deep representation learning to properly initialize network on huge healthy patient cohorts and then transfer them to study clinical pathology of interest in smaller cohorts. On the other hand, we have identified multi-view variational auto encoders as good candidates to integrate multiple modalities. Moreover, we challenge the common assumption that neural networks are not interpretable. We use a digital avatar procedure as an interpretability module capable of reporting the inter-view relationships learned within a multi-view autoencoder. In particular, we integrate this procedure into a novel framework that combines multiple interpretations and utilizes stability selection to identify meaningful and reproducible associations between brain-imaging modalities and behaviour. We apply this framework to exhibit specific brain-behaviour associations present in the transdiagnostic cohort Healthy Brain Network (HBN). The identified brain-behaviour associations establish connections between regional cortical features from structural magnetic resonance imaging and electronic clinical record forms assessing psychiatric symptoms. We show this framework is able to find relevant and stable associations.

Remerciements

Je souhaite en premier lieu remercier mes encadrants, Vincent et Antoine, pour leur accompagnement exemplaire. En tant que doctorant, j'ai pu entrevoir les difficultés que peut amener ce travail de longue haleine, parfois solitaire. Les encadrants ont un rôle crucial à y jouer et ne devraient jamais le négliger, au risque de créer de la rancœur dans les meilleurs cas, des issues plus difficiles dans d'autres. J'ai une telle chance d'avoir pu être encadré par deux personnes qui ont réellement à cœur que ma thèse se déroule au mieux et qui y ont investi beaucoup de temps et d'énergie. Pour tout ça, merci. Merci Vincent pour ton soutien et ta facilitation sur cette dernière année de thèse. Je suis toujours enjoué à l'idée de discuter de nos travaux avec toi. Merci pour ta sollicitude et de ne pas hésiter à poser les bonne questions qui permettent de garder le cap. Merci pour ton aide et ton soin particulier à structurer, rédiger et relire nos productions scientifiques. Merci Antoine pour ton soutien technique et ta disponibilité. Merci pour ton soin à rendre notre contenu méthodologique disponible avec des initiatives pour créer des espaces de partage scientifique. Merci aussi pour ton aide à rédiger, relire et proposer des figures pour égayer nos articles.

Ensuite, j'aimerais remercier mon jury. Tout d'abord, merci Diana Mateus pour ton examen très soigné et approfondi de mon manuscrit et ton rapport élogieux, ainsi que pour tes suggestions qui ont permis de l'améliorer. Thank you Professor Mourao Miranda for accepting to review this thesis, your careful examination of the manuscript and kind comments and questions. Merci Professeure Allassonnière d'avoir accepté de participer et présidé ce jury. Merci Charles Laidi d'avoir accepté d'évaluer ma thèse et su apprécier mes contributions applicatives et méthodologiques malgré le fossé entre nos expertises. Merci à tous pour nos discussions enrichissantes et vos questions stimulantes.

Je souhaite aussi remercier toutes les personnes de mon école doctorale EOBE ou de l'université qui ont permis de s'assurer du bon déroulement de cette thèse. En particulier merci à Jean-Cristophe Ginéfri qui a accepté de fonctionner avec des délais de fin de thèse réduit et a toujours été attentif et disponible pour moi. Merci à Emilia Davodeau, Melissa Starek et Silvie Pommier, ainsi qu'à la Maison du Doctorat.
De la même façon, je souhaite remercier Pietro Gori et Marco Lorenzi pour leurs interventions dans mon comité de suivi, qui m'ont permis d'avoir un regard extérieur sur mon travail de thèse et sur l'encadrement, ainsi que des suggestions intéressantes.

Ensuite, je voudrais remercier Neurospin et le CEA pour son accueil et ses conditions de travail très confortables, ainsi que l'attention qu'il y est porté pour que celles-ci le soient pour toutes et tous. Je souhaite particulièrement remercier Maryline Hévin pour son accompagnement pour les missions, impressions, et autres tâches administratives ponctuelles, ainsi que son travail pour informer les salariés sur leurs droits et opportunités liées au comité d'entreprise. Je souhaite aussi remercier tout le personnel du bâtiment, Fabrice Poupon, anciennement Joël Cotton, le personnel de l'accueil et les agent.e.s d'entretien qui veillent à ce qu'on puisse travailler dans un cadre agréable. Merci

également aux personnes gérant les associations pour les salarié.e.s et aux bibliothécaires du CEA.

Maintenant, j'aimerais remercier mes collègues et collaborateurs. Vous avez rendu ces 4 années très sympathiques et je suis heureux d'avoir pu discuter, travailler, rigoler et tant d'autres choses avec vous. Merci aux coureurs du vendredi Nico, Edouard, Thibaut, Aymeric, Joël, Robin, Benoît, Josselin, Julien, Julien, François, Raphaël pour ces sorties par tout temps qui ont rythmé mes semaines, ces repas agréables, et merci à François pour les sandwichs.
Merci au groupe de lecture sur le deep learning que j'ai eu le plaisir de gérer pendant une belle année, merci aux personnes à l'initiative du groupe Benoît, Louise et Zaccharie, et merci à Pierre et Sara d'avoir pris le relais, ainsi qu'à tous ceux qui ont pu participer et contribuer pour enrichir notre connaissance de ce domaine passionnant en transformation perpétuelle.
Merci à Jean-François pour ton intérêt scientifique, merci à Cathy pour tes conseils et ta gentillesse, merci Edouard pour ta bonne humeur et ta jovialité, merci Nico pour ton intérêt pour les autres, merci Clément de vouloir élaborer une vie de laboratoire, merci Sara pour tes blagues que je ne comprends pas toujours, merci Pierre pour ta bienveillance et d'être toujours prêt à rendre service. Merci Robin pour nos discussions scientifiques et autres autour d'un café, merci Chloé pour nos pauses-complaintes de galères de fin de thèse, nos aventures de voyage et pour nos séances d'escalade, merci à Loïc aussi. Merci Pierre-Yves d'être toujours prêt à discuter quand on ne veut pas travailler, merci Aurélie pour le voyage en Colombie. Merci Ulysse de m'avoir permis de partir en week-end pour me détendre avant ma soutenance en me ramenant mes clefs. Merci à Ophélie et à Louise pour nos retours en transport du laboratoire, qui permettent de partager et de se changer les idées en fin de journée. Merci Thomas pour nos séances d'escalade et nos pauses-balades autour de Neurospin. Merci Clara pour tes tuyaux et d'être toujours là au bar, même avec tes petits. Merci Julie pour ta gentillesse. Merci à Cyril pour ton entrain et tes clowneries qui nous ont largement soudés. Merci Benito (alias le Beun), toi qui m'a montré la voie vers la recherche et le haut du mur d'escalade. Merci pour nos discussions matinales dans le bus quand je ne te le faisais pas rater, merci pour tes relectures et ton aide pour nos publications. Merci à tous d'avoir été présents pendant ma thèse, au laboratoire comme au bar.

Évidemment, je ne peux pas oublier de remercier mes amis, qui sont pour moi une source d'inspiration et de soutien infini. La limite est fine entre collègues et amis et il est évident que certaines personnes déjà mentionnées appartiennent à cette catégorie et se reconnaîtront. Thomas, merci pour ces trois belles années de colocation, durant lesquelles nous avons partagé bien plus qu'un logement. Merci pour ton soutien et merci de me montrer la richesse des relations. Hélène prend le relais mais je sais qu'on saura se retrouver sur deux roues, sur un mur, derrière un pc ou autour d'un bon repas. Merci Kimou de me faire partager toutes tes aventures sportives et les autres, merci pour ton courage et ta force qui m'inspirent toujours. Merci Olivia pour ta bonne humeur et ta gentillesse. J'ai la chance de compter quelqu'un d'aussi bienveillante et attentionnée que toi parmi mes amis. Merci Oriane d'être là dans les moments difficiles aussi, merci Luis d'avoir partagé toutes ces belles années d'études plus ou moins studieuses. Merci Adrien pour tes petites attentions qui font toujours plaisir et nos vacances de qualité, merci Antonin pour ces moments sportifs partagés à Massy comme à la montagne en toute saison. Merci Hélène pour ta confiance et ces moments de

partage, merci Javita pour nos séances d'escalade et nos aventures franco-suisses. Merci Erwan pour ta gentillesse et pour nos voyages, merci à Simon et Matthieu. Merci Ambre, Erine et Arthur pour nos moments d'évasion, que ce soit en vacances ou dans un escape game. Mathieu et Elisa, alias les voisins, merci pour cette cohabitation et tous ces moments, sportifs et autres, partagés. Et enfin, merci Nana pour ton soutien et ta tendresse. Il est évident que cette année aurait été difficile sans toi. Tu m'as aidé à surmonter mes doutes et mes angoisses, même la veille de ma soutenance. Merci de partager ta vie avec moi.

Finalement, je souhaite également remercier ma famille. Papa, Maman, je ne pourrais jamais vous remercier assez pour tout ce que vous avez fait pour moi. Merci de m'avoir offert un cadre favorisant pour grandir, apprendre et me développer vers ce que je suis aujourd'hui. Il semblerait que j'ai du mal à trop m'en éloigner… Je souhaite aussi remercier Grand-Père, qui m'a appris et montré ce que peuvent nous apporter la forêt et son sol mycélaire, Marou qui m'a enseigné le soin pour les autres. Merci Gaëlle et Patrick d'être là pour célébrer la famille, merci à mes cousins Paul, Mathieu et Armand d'avoir partagé tous ces moments. Merci Papi et Mamie pour nous avoir fait découvrir la vie en ville comme à la campagne. Aurélien, merci pour ton soutien pendant ces trois dernières années et merci de partager tes enfances avec moi, la tienne comme celle de ta fille.

Contents

Acronyms

K-NN K-Nearest neighbours. 33, 43, 60

AD Alzheimer's disease. 18, 22, 25, 56, 81

ADHD Attention Deficit Hyperactivity Disorder. 25, 73, 97, 98, 114, 115, 118, 124

ADNI Alzheimer's Disease Neuroimgaging Initiative. 25

AE Auto-Encoder. 42

ANN Artificial Neural Network. 9, 36

ASD Autism Spectrum Disorder. 21, 73, 97, 98, 100, 102, 104–106, 114, 116, 118, 124

BAcc Balanced Accuracy. 62–66

BD Bipolar Disorder. 24, 25

CCA Canonical Correlation Analysis. 10, 75–77, 82, 91, 121, 148

CNN Convolution Neural Network. 36, 38, 39, 42, 55

CNV Copy Number Variation. 23, 74, 122

CPU Central Processing Unit. 68

CSF CerebroSpinal Fluid. 17, 21, 23, 25, 51, 55

DA Digital Avatar. 10, 84–87, 91, 112, 119, 120

DAA Digital Avatar Analysis. 10, 11, 83, 85, 87–90, 92, 105, 106, 108, 110, 118–120, 124, 148–150

DiNe Direct Neighbour. 54–57, 59, 62, 66, 68, 125, 145

DL Deep Learning. 11, 16, 26, 36, 40, 41, 55, 56, 72, 82, 105, 116, 125

dMRI diffusion MRI. 17, 118

DNA DesoxyriboNucleic Acid. 23–25

DSM Diagnostic and Statistical Manual of Mental Disorders. 9, 72, 73, 94, 96, 104, 117, 127, 143, 147

eCRF electronic Clinical Record Form. 106–109, 112–114

EEG Electroencephalography. 22, 26, 74

ELBO Evidence Lower Bound. 78, 79

ERM Empirical Risk Minisation. 32

ERP EquiRectangular Projection. 53, 66

fMRI functional MRI. 17, 21, 26, 76, 118, 126

FSIQ Full Scale Intelligence Quotient. 55, 62–64, 97–99

GAN Generative Adversarial Networks. 39, 43

GLM Generalised Linear Model. 33

GPU Graphical Processing Units. 36

HBN Healthy Brain Network. 11, 47, 62–66, 94, 96–99, 104–108, 115, 116, 119, 121, 124, 127, 147, 149

MAE Mean Absolute Error. 62–66

MCMC Monte Carlo Markov Chain. 77, 81

MDS MultiDimensional Scaling. 35

MEG Magnetoencephalography. 22, 26, 74

ML Machine Learning. 32, 37, 41, 56, 74, 92

MLP MultiLayer Perceptron. 36–38, 42, 57, 62, 63, 104, 107, 145

MoPoE Mixture of Product of Experts. 79, 80, 106–109, 112, 117, 118, 120, 126

MRI Magnetic Resonance Imaging. 16–18, 20, 22–27, 32, 50–52, 54, 59, 61, 69, 72, 74, 81, 97, 107, 117, 118, 121, 124, 126, 127

mRNA messenger RiboNucleic Acide. 24

mVAE multi-view Variational Auto-Encoder. 10, 11, 77, 78, 81–89, 91, 92, 94, 97, 99, 104, 105, 116, 121, 124–127, 148–150

NIMH National Institute of Mental Health. 73, 94

NN Neural Network. 9–11, 36–38, 40–44, 47, 52, 54–56, 66–69, 72, 75, 77, 82–84, 104, 105, 120, 123–126

OpenBHB Open Brain Healthy Bank. 27, 56, 62–66, 127

PCA Principal Component Analysis. 35, 60, 76, 82, 102, 126

PET Positron Emission Tomography. 22, 25, 81

PGM Probabilistic Graphical Model. 47, 72, 75, 77, 78, 81, 91, 148

PLS Partial Least Squares. 75, 76, 82, 91, 121, 148

PoE Product of Experts. 79, 81, 100, 126

QC Quality Control. 62, 97

r-DAA regularised Digital Avatar Analysis. 10, 11, 87–90, 92, 105, 106, 108, 110, 111, 120, 124, 148–150

RDoC Research Domain Criteria. 9, 10, 73, 91, 94, 95, 117, 124, 143, 147, 149

ReLU Rectified Linear Unit. 39, 107

RePa Rectangular Patches. 53, 55, 66

RNN Recurrent Neural Network. 40

ROI Region of Interest. 10, 21, 52, 68, 99, 102–116, 119, 124, 125

RSA Representational Similarity Analysis. 89, 108, 110, 111

SBM Surface based Morphometry. 18

SCNN Spherical Convolutional Neural Network. 9, 10, 57, 61–66, 68, 69, 120, 123, 125, 145, 146

SDQ Strengths and Difficulties Questionnaire. 97, 106, 109, 111–116, 118, 119

SGD Stochastic Gradient Descent. 36, 38

sMCVAE sparse Multi-Channel VAE. 81, 97, 99, 101, 102, 104, 116, 121, 126, 149, 150

sMRI structural MRI. 17, 18, 23, 50, 51, 55, 72, 74, 123

SNP Single Nucleotide Polymorphism. 23, 25, 26, 74, 76, 122

SotA State-of-the-Art. 40, 81, 84

SRS Social Responsiveness Scale. 97, 98, 104, 106, 109–116, 118, 119, 149

SSL Self-Supervised Learning. 9–11, 42–44, 56–58, 61, 65, 67, 69, 72, 123, 125, 126

TCGA The Cancer Genome Atlas. 25, 76, 81

TL Transfer Learning. 43

VAE Variational Auto-Encoder. 39, 68, 77, 79–84, 100, 105–109, 112, 117, 118, 120, 126

VBM Voxel based Morphometry. 18, 27, 127

Introduction

Mental health is a major issue as 970 millions people were living with a mental disorder in 2019, according to the World Health Organisation (WHO)[1]. However, these diseases are often overlooked and no sufficient resources are invested for their treatment. This is partly due to the fact that they are rather poorly understood conditions. The recent development of non-invasive and high throughput systems to get insights about biological systems and their functioning has the potential to change this. But other paradigms should change to allow for a global knowledge progression regarding brain pathology. For psychiatric disorders, the conventional approach was to consider a single diagnosis and condition and use a statistical model to predict it using a single biological assessment such as an imaging modality. However, it has been acknowledged that results obtained with this strategy are unsatisfying and do not allow knowledge to progress in psychiatry. The major hypothesis for this issue is the recognised heterogeneity of such categorical conditions as defined by classification tools such as the Diagnostic and Statistical Manual of Mental Disorders (DSM). The phenomenon of comorbidity between conditions and interactions between biological processes, which can be captured at different levels, are overlooked by these diagnosis based studies. Other initiatives such as the Research Domain Criteria (RDoC) advocate for a more comprehensive approach to studying psychiatric disorders by incorporating diverse data that reflects different level of the organisational complexity of life (e.g. imaging, genetic and symptoms). The RDoC principles suggest that a thorough description of a pathology requires consideration of multiple dimensions that may be shared across different psychiatric syndromes and may even contribute to non-pathological variability. In line with these recommendations, this thesis work aims at providing tools compatible with such criteria, and in particular tools relying on Artificial Neural Network (ANN). These tools should be able to maximally model structure consistent with a priori knowledge about biological components. Population imaging studies designed to investigate psychiatric disorders often include different assessments, which inform about different biological components.

In this context, we first propose to model biological cortical structure using the spherical representation obtained with the surface based morphometry pipeline from FreeSurfer. Specific NN operators have been developed to account for this particular mesh structure when learning from such cortical surface data. In particular, we propose a Self-Supervised Learning (SSL) pipeline that enables such Spherical Convolutional Neural Network (SCNN) to learn from cortical surface data. This pipeline relies on carefully designed data augmentations which implements modifications of the cortical measures over the spherical mesh. Among these 5 proposed augmentations, 3 of them adapt classically applied to natural images augmentations to the spherical structure. The 2 other augmentations relies on biologically plausible assumptions, which propose to introduce some amount of noise in the cortical measurements coming from either other similar participants or the other subject's hemisphere. These augmentations allow the SCNN to learn representations containing a significant amount of semantic information and directly applicable to other cohorts with

[1]Mental Health WHO's topic.

good generalisation properties. This is the first work which proposes tools to maximally leverage cortical surface structure in order to teach SCNNs in an unsupervised manner. This work has interesting applications in large general population imaging without diagnostic, which allows to use such SSL scheme to properly initialise these SCNNs before applying them in clinical cohorts with scarcer data.

In line with directions set by the RDoC, we attempt to maximally integrate available biological knowledge in the learning algorithms. We describe issues that come with integrating multiple assessments and discuss the available tools to overcome them. Linear integration using Canonical Correlation Analysis (CCA) or assimilated display interesting characteristics, such as the ability to integrate an arbitrary number of modalities and their interpretations capabilities. However, they lack of capacity due to their linear formulation, preventing them from directly exploiting biological structure, and can not properly handle examples with missing assessments. Probabilistic graphical models, and particularly multi-view Variational Auto-Encoders (mVAEs) stand out for multiple reasons. First, their formulation allow them to leverage NNs flexibility and instantiate different adapted encoder and decoder for each modality (e.g. SCNN for cortical surface data). Second, they are naturally able to use data with missing assessments. Last, they display disentanglement properties, in that when modelling modality-specific latent spaces additionally the joint latent space, shared sources of variability are captured by the joint latent space, and specific sources remain in their modality's specific latent space. This allows to properly handle imaging acquisition site effects, which can be very harmful in population imaging, without additional harmonisation technique.
On the other hand, relying on NNs prevent them from being naturally explainable. Nevertheless, key properties of mVAEs allowed us to design an interpretation module, relying on realistic perturbations of features from one modality, and monitoring their impact on the generated features from other modalities. These altered data examples are called Digital Avatars (DAs) and are further analysed using hierarchical linear models, producing association coefficients between each perturbed feature and every features from other views. This whole interpretation framework is called Digital Avatar Analysis (DAA). Of course, such interpretations are not exempt of the various sources of variability impacting the solutions learned by NNs. Aleatoric uncertainty is due to unreducible population variability, while epistemic uncertainty arise from modeling choices. Epistemic uncertainty can be significantly larger in NNs, considering the same architecture and hyperparameters, than in linear models. Provided a train / left-out split of a dataset, we propose a regularised version of the DAA (r-DAA), which consist in repeating n_E times: training the mVAE on the training set then interpret the trained mVAE using the left-out subjects. Only the random initialisation and batches during training vary. Equipped with a proper ensembling function, these associations are aggregated and discretised, providing results robust to epistemic variability. The aleatoric variability is controlled by using a stability selection procedure, which consists in retaining associations consistently selected across numerous train / left-out splits of the population.
In particular, we display the ability of this whole pipeline to output reasonable brain-behaviour associations in a transdiagnostic cohort of children presenting behavioural peculiarities. We integrate clinical scores assessing expressed symptoms with regard to cortical measures over Regions of Interest (ROIs) of the cortex, instantiating all the previously introduced tools.

Thesis organisation

This thesis presents mostly methodological work. However, we provide intuitions about possible applications in the study of psychiatric disorders.
In the first part, which regroups Chapter 1 and 2, we want to introduce different notions. Chapter 1 presents brain magnetic resonance imaging, describing the brain cortical anatomy, and population imaging, which often regroups multiple assessments for each participant and introduces novel challenges. In Chapter 2 we give a brief introduction to basic statistical learning ideas, with a few examples, and describe more thoroughly Deep Learning (DL) and associated concepts.
In the second part, we present our contributions. Each chapter of this part provides a short state-of-the-art related to the addressed research questions. Then it describes the contributions and associated experiments. In Chapter 3, we try to address the question of modeling biological cortical structure using an adapted NN. Our contribution is to propose multiple cortical surface data augmentations and assess there relevance in a SSL scheme. Chapter 4 describes the heterogeneity problem in psychiatry and discuss different integration methods. Then it introduces the DAA interpretation module to alleviate the lack of explaibility in mVAEs. It also proposes the r-DAA and stability selection procedures design to produce stable associations. Note that Chapter 4 does not contain experiments, as validation of the introduced methodological tools is propagated in the next chapter. And Chapter 5 introduces the concept of transdiagnostic psychiatry, then presents different studies of the cohort Healthy Brain Network (HBN). The first focuses on stratifying autistic subgroups in the cohort, where the second applies the introduced framework in Chapter 4 to discover reproducible brain-behaviour associations, displayed in a transdiagnostic perspective.
Finally, we conclude the thesis by providing a summary of the explored problematics, proposed solutions and their potential impact and perspectives.

Part I

Background

1 - Brain magnetic resonance imaging

Contents

In this chapter, we introduce an imaging technique to inspect in-vivo organs widely used today. Magnetic Resonance Imaging (MRI) allows to visualise precisely in-vivo soft tissues up to 0.2mm isotropic with the most precise human MRI machine, such as the prototype recently installed at NeuroSpin, which features a record magnetic field of 11.7T. The resolution of MRI is a question of time and magnetic field strength. With a sufficiently long acquisition, lower field MRI machine could attain this resolution. In practice, this is impossible due to human movements that would inevitably produce blurred images. The quest for new biomarkers does not necessarily depend on extreme acquisition parameters. First, biomarkers produced at 11.7T would not necessarily be detectable at lower resolution with usual MRI machines deployed in hospitals for instance, which are most often equipped with 1.5T or 3T MRI machines. Second, such a powerful machine will never be deployed to multiple centres, due to its physical characteristics: it weights 132.000Kg, requires a cumbersome cooling system and costs 70M€ [38]. Therefore, expected biomarkers at 11.7T are not being searched for condition diagnostic or prognostic, but rather as tools to understand intrinsic mechanisms of the brain, related pathologies and human cognition.

An alternative to investigate brain characteristics and disorders is to collect large compendiums of data in order to unveil biomarkers of interest. This approach also allows to better take into account the specific characteristics of each population. Nowadays, huge research projects have built shared and open resources from health data in large population, and other initiatives exist, such as in population imaging, which promote high-quality phenotyping with imaging and psychological scoring, while specifically recruiting very large numbers of people into a research protocol. These initiative are in line with the growing need of data for recent statistical models such as Deep Learning. It has also been recently shown that discovering reproducible associations between brain and complex behavioural phenotypes requires thousands of individuals [240]. However, collecting a large set of data requires regrouping data from multiple sources and imaging acquisition sites. This poses new challenges, such as site effect harmonisation, which is mandatory to ensure downstream modelling is not impacted by this non-biological variability [136, 357].

In the first part of this chapter, we will introduce MRI for brain study. We will then describe the brain anatomy, focusing its outermost layer, the cortex. In the second part, we will try to introduce population imaging and describe its multimodal aspects. Then we will shortly introduce a few population studies and finally present the main challenges we identify when analysing and building tools from such studies for clinical application.

1.1 . Brain anatomy inspection using MRI

Magnetic resonance imaging relies on magnetic properties of water molecule contained in our body, and more precisely on the magnetic properties of hydrogen atom nuclei, the proton. A proton is rotating on itself and is characterised by its intrinsic angular momentum or spin. When subject to a strong magnetic field, hydrogen spins align according to the field's direction. The MRI machine perturbs the spins using radio wave in another direction which makes them move out of this

[1]Link to the web page.

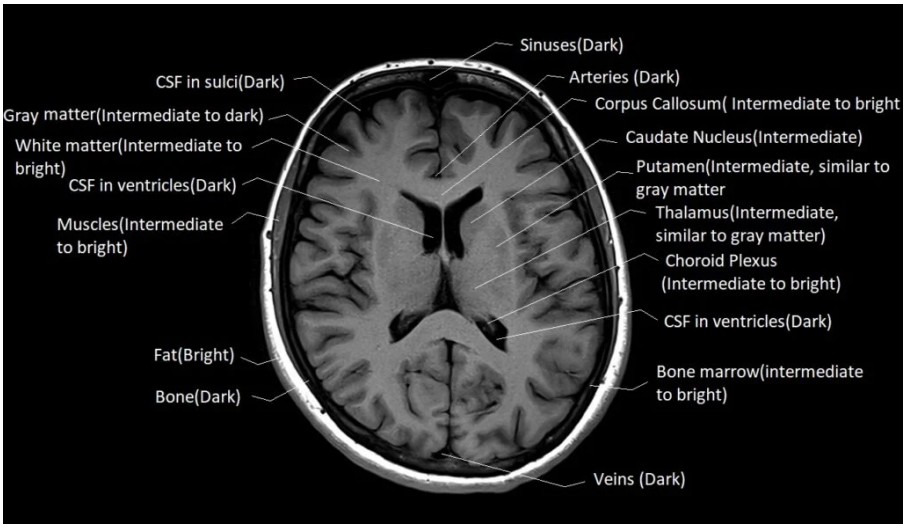


Figure 1.1: Axial view of T1-weighted brain MRI with annotations. Image found from an external source[1].

original position. When these waves are turned off, the spins progressively return to their original position and send back a radio signals. The spins go back to their initial position at different speeds, depending on the tissue they compose. Major advantages of this technique compared to other imaging methods is that it can produce 3D images, with the same resolution in all direction without exposing the patient to harmful radiation. The magnetic resonance imaging technique efficiently captures various tissue natures in different contrasts.

From the same principle, multiple imaging types can be obtained, by adapting MRI sequences in order to highlight various tissues with different contrasts (e.g. T1-weighted vs T2-weighted) or local micro-structure properties which underlie tissue organisation such as white fibber bundle with diffusion MRI (dMRI). Specific sequences are sensitive to local oxygenation level of haemoglobin and reveal dynamic neuronal activity using Blood Oxygenation Level Dependent (BOLD) functional MRI (fMRI). In this section, we will focus on presenting brain structural MRI (sMRI), the most basic MRI technique which consists in inspecting brain tissues and therefore the related anatomy.

1.1.1 . Structural MRI

Structural MRI is used to inspect soft tissues. When dealing with brain MRI, we are able to visualise and identify the different parts of the brain, grey matter with the different well-known cortices and sub-cortical components (e.g. thalamus and deep nuclei), as well as white matter and ventricles. We can see the head boundaries and the skull shape, and the CerebroSpinal Fluid (CSF) whose role is to maintain the brain isolated from the skull and transport chemicals in the ventricles. These components can be visualised in the T1-weighted MRI provided in Fig. 1.1. The grey matter appears darker than the white matter in this contrast.

Applications in clinical routine. So far, MRI has been used in clinical routine by radiologist to help the neurologist to diagnose diseases where biomarkers are visible by the human expert eye, such as brain tumours or Alzheimer's disease (AD) atrophy in the hippocampal region. MRI mostly inspects tissues and is thus not pertinent to spot bone defaults or fractures, because of their low concentration in water. It is used for treatment monitoring as well. Some decision aiding radiological tools start to emerge from MRI research but they are still confined to diseases that manifest themselves through spatially extended phenomena[95], and many brain diseases, specifically psychiatric disorders, remain poorly known with very few verified biomarkers that can not be found using expert eyes. For the investigations on brain diseases, popular MRI studies based on healthcare data include oncology [322, 64], neuro degenerative [380, 81] (e.g. AD and Parkinson's disease) and psychiatric disorders [122, 112, 1, 377]. These studies recommend to use multimodal biomarkers, including genetic and molecular assessments, additionally to imaging biomarkers. However, studies in clinical routine requires for them to be robust to variation in acquisition settings and population biases, which is very hard to come by in practice because most research cohorts are built with only a few acquisition sites. Even when data are collected through multiple sites, multiple MRI constructors and across multiple population, harmonisation or site effect handling is not properly done, as it is still an open research question (see Section 1.2.3 for further information). Validation of biomarker detecting tools in clinical settings is crucial for translating research to clinical routine.

Research applications. Other studies focus on research data with subjects specifically recruited and submitted to advanced imaging protocols. These studies are designed to find biological correlates of diseases using imaging biomarkers, not aiming to be used in clinical routine, but rather uncovering hidden biological machinery that can help us better comprehend brain disorders. Understanding these mechanisms can in turn lead to design specific treatments to modify such abnormal functioning. Structural markers extracted using MRI depend on the studied pathology. For instance, oncology will focus on studying the tumour tissue, as compared with healthy tissue, often computing quantitative measures characterising local properties such as radiomics [322]. Studying neurodevelopmental or neuropsychological disorders will often rely on heavy expert preprocessing tools to extract relevant measures from the 3D volumes such as cortical thickness. This cortical thickness is analysed locally or integrated over some regions of interest, defined by an anatomical or functional atlas, which allows results interpretation [34, 223]. These tools either use Voxel based Morphometry (VBM) [131] preprocessing to extract and register the 3D volume in a template space and conserving solely relevant tissues, or Surface based Morphometry (SBM)[86, 120], which concentrates on the cortex and can extract various proxies of its geometrical shape.

1.1.2 . Brain cortical anatomy

In particular, sMRI allows us to visualise and describe the main cortical areas of the brain. The brain is composed of two almost symmetrical hemispheres, with mostly similar functions. Neuroanatomists have been able to assign specific functions to the different parts of the cortical surface, referred to as cortices (or lobes for larger parts), illustrated in Fig. 1.2 and further described below:

- **Frontal cortex** The frontal cortex (or frontal lobe) regroups the largest part of the cortical sur-

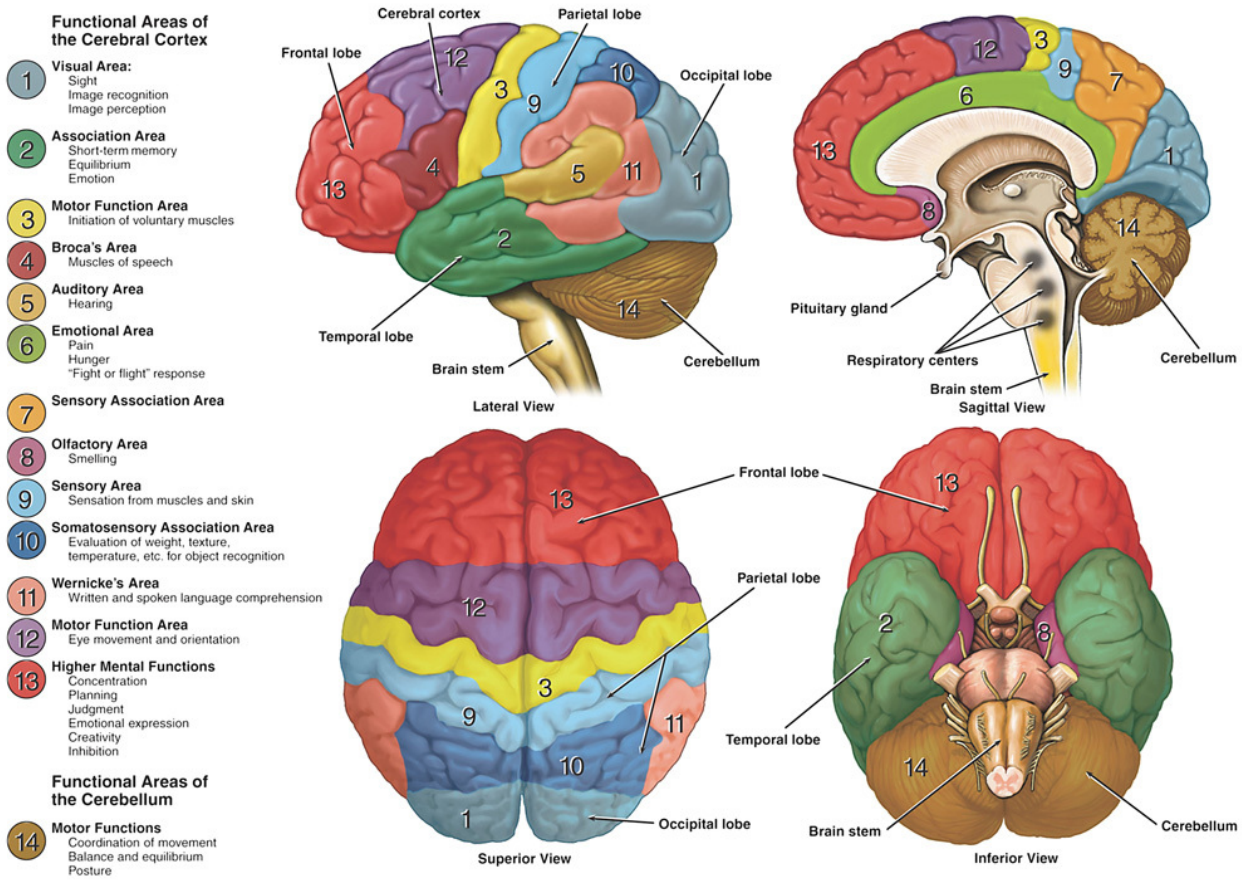


Figure 1.2: Illustration of the human brain anatomy and functions. This image was drawn from a web article[2].

face, and is certainly the most complex. It handles a wide variety of processing related to executive function, reasoning, decision-making, speech with the Broca's area (left hemisphere mostly) and movement execution.

- **Motor cortex** The motor cortex is part of the frontal lobe and delivers signals for motor initiation (premotor), execution (primary motor) and coordination (supplementary motor area). Different parts of the cortex command different parts of the body and have been mapped together, and the left part of the brain controls for right part of the body, and inversely, for most people.

- **Parietal cortex** The parietal cortex (or lobe) is situated at the upper back part of the brain. It is called the somatosensory cortex as it is known to be responsible for sensation and touch information processing, as well as spatial positioning. The Wernicke's area is part of this cortex and is implicated in text and speech understanding. The partietal cortex is known as well for integrating sensory and motor information as input for the premotor cortex.

- **Temporal cortex** The temporal cortex (or temporal lobe) plays a central role in auditory in-

formation processing. It is thus implicated in speech recognition as well and its posterior part is connected to the parietal lobe through the Wernicke's area. It is also linked to movement perception and some memory related functions such as face recognition.

- **Occipital cortex** The occipital cortex (or occipital lobe) is located at the most lower back part of the brain. It is mostly responsible for processing visual information and interpreting it, rooted in the primary visual cortex.
- **Limbic cortex** The limbic cortex is not generally associated with a specific cortex lobe, but is grouped within the limbic system, which includes cingulate cortex areas surrounding the corpus callosum, the hippocampus and the amigdala. This complex system plays a key role in emotion processing, behaviour and long-term memory. In particular, the limbic cortex is responsible for modulating emotion, visceral motor processes and plays an associative role between the hippocampus and other cortical areas.
- **Insular cortex** The insular cortex is the dark side of the brain: you can not visualise it in Fig. 1.2. It is hidden inside the lateral fissure separating the temporal lobe from the frontal and parietal lobes, buried between the folds. Its role is less known than other cortices, but it plays a role in integrating multiple types of information related to emotional processing, including taste, pain, hunger, sensation, social functions such as empathy and vestibular (related to self-awareness perception and inner ear) functions.

The brain cortex compose the outermost part of the brain. It represents most of the grey matter of the brain, which are the tissues regrouping most of the neuronal cells. Neuron activity presents itself as an electrical and a chemical signal travelling across the brain and between two cortex areas or between cortex, cerebellum and subcortical components regrouping notably the thalamus, hypothalamus, hippocampus, putamen, basal ganglia and deep nuclei. As the cortex commands most human abilities, studying it can give valuable insights into the inner cogs of the brain, to better understand brain function and dysfunction. Its highly folded structure allows it to drastically increase its area and grey matter volume, but makes its study challenging, due to its particular shape.

1.1.3 . Cortical surface study

The brain cortical surface organisation is constantly the subject of new discoveries and scientists have been trying to study it for a long time now. Recent biological imaging tools such as microscopic imaging used in histology and MRI allow to improve its observation at different scales and thus deepen our understanding of it. For instance, we know that neurons in the cortex have a laminar organisation, i.e. are built in separate layers with different properties, aggregated to form the cortex, and that neurons are organised in perpendicular columns with respect to the pial. Histology has provided insights into the composition of these different layers, which in turn can help us better understand the formation of the cortex by driving hypotheses, its folding mechanisms and what can be considered as normal or abnormal development. In a complementary way, MRI allows us to inspect the brain and its cortex at a macro scale, from which we can measure characteristics and qualify them. MRI is a non-invasive in-vivo imaging technique while histology requires to extract

[2]You can find this article here.

tissue, so it usually screens dead or abnormal tissue obtained from ex-vivo brain or biopsies.

Folding patterns. The mechanisms responsible for the cortex folding are still unknown, although its study lead to various hypotheses. The folding patterns, defined by the gyri (bumps of the cortex corresponding to outer folds) and sulci (parts of the inner fold filled with CSF), can be compared to fingerprints: they are unique to each individuals, although some of them are found in all healthy human brains. These patterns have allowed to build sulci atlases [281] and have been linked to some brain disorders (e.g. the "Power Button Sign" pattern of the central sulcus for drug resistant epilespy [251]). Their study is challenging due to their high variability across individuals [239]. Their origins is partly attributed to genetic causes [217], and in particular folding patterns from twins are more similar to each other than with other brains folding. Other hypotheses try to explain folding pattern using mechanistic models, thus modelling various opposing strengths underlying biological processes of the cortex development [118]. Cortex folding happening during the last three months of pregnancy makes it difficult to monitor in longitudinal studies. It can only be studied in this condition where the mother limits imacting environmental factors [20].

Geometrical properties of the cortical surface. In parallel to studying the cortex using folding patterns, we can also consider various measures of the cortical surface. Expert tools such as FreeSurfer [86, 120] have been developed specifically to extract geometrical measures of the cortex shape, such as cortical thickness, surface area, curvature or sulcal depth, and represent them on suited discretised geometric structures. They rely on segmenting the grey matter borders, tessellating the grey / white interface and computing these measures. The resulting data can be projected on a common representation space which preserves the cortical surface biological properties. These techniques will be further described in Section 3.1. In subsequent analyses, these measures can be either analysed directly in such representation, or aggregated on ROIs defined by an atlas of the cortical surface.

1.2 . Population imaging

For research purposes, consortiums have been built around common purposes, e.g. finding anatomical biomarkers for Autism Spectrum Disorder (ASD). These efforts have allowed the blooming of large scale cohorts with specific objectives. Population imaging describes the acquisition of medical images in such controlled population. Their acquisition setting and recruitment protocols are completely driven by these objectives. As a result, the number of cohorts available for a specific application or research question is limited by the occurrence of a pathology in a population for instance, and correlates to the interest aroused by the research question.

For cognitive neuroscience, specific task-based imaging such as fMRI techniques are used within precisely designed and conducted protocols, limiting the experiments to a single acquisition setting and a limited number of participants ($\sim$ 5-200). On the other side of the spectra are oncology, psychi-

atric diseases and general purpose research cohorts. They often regroup larger spanned populations, regrouping imaging from multiple acquisition sites and include healthy subjects. They can be composed of a population ranging from a few hundreds to tens of thousand subjects.

1.2.1 . Multiple assessments

These population imaging cohorts are often richly phenotyped and include several types of assessment (i.e. ensemble of measures using different techniques and targeting several biological aspects). They always include at least two types of information: a few modalities of imaging and participant related information such as age and sex, mandatory for most research analyses. Now, most cohorts include other types of assessment, such as different modalities of imaging, genetic, clinically relevant factors such as diagnosis or symptoms related information, and other metadata.

Imaging modalities. There is a wide panel of imaging modalities that can be available in brain disorder population study. Each one of them is made to shed light on different properties of the brain, which can be evolving at different time scales and are visible at different spatial scales.

For instance, brain activity changes rapidly (100 to 200ms on average), very related to human activity, and tools to measure it need to be able to measure this activity with a sufficiently high frequency to capture such changes. Functional MRI measures neuronal activity using as proxy the concentration of deoxyhemoglobin, which results from neuron oxygen consumption during activation, and thus increases locally when neurons activate. This is called the BOLD effect and can be measured spatially at MRI precision level, between 1 and 2mm isotropic, depending on the magnetic field. On the other hand, other techniques exist to capture temporal signal with higher frequency, namely electro–or magneto–encephalography (respectively EEG or MEG). They rely on electrodes capturing respectively electric or magnetic signals stemming from brain activity from outside of the brain. They can be considered as imaging since they enable to reconstruct map of the electromagnetic field. Yet, they have a very poor spatial precision, limited by the size of the electrodes and their distance, because we don't want them to capture the same signal. This imaging technique is also very noisy due to many surrounding factors to which this technique is sensible, such as the current frequency, which is different between different countries or continents, or patient irreducible variations such as heartbeat or blinking. Another technique called ElectroCorticoGraphy (ECoG) captures the same signal as EEG, but directly on the brain cortical pial. It is thus a very invasive technique allowed in Europe in human patients suffering of very specific diseases such as intractable epilepsy. It provides a signal with a greater Signal-to-Noise Ratio (SNR) than EEG because it is not hampered by the skull and other head layers.

Changes in metabolic or physiological processes, usually happening at a slower speed than neuronal activity, can be captured using Positron Emission Tomography (PET). It uses different radioactive tracers suited to capture the targeted process. As a commonly used tracer, Fluorodeoxyglucose (FDG) PET is often used to visualise brain tumours which alter glucose metabolism and ^{11}C-PiB for AD to image amyloid plaques.

For assessing tissue evolving at a slower pace, such as macro scale organisation of a tumour, or brain anatomy, the use of sMRI is preferred, with different weightings, depending on what we want to inspect. T1 weighting give a higher contrast to fatty tissues, such as the myelinated axons in the white matter, and suppresses the signal of water, where T2 weighting enhances the signal of water. Fluid-Attenuated Inversion Recovery (FLAIR) contrast allows to suppress fluid signal, made to suppress the CSF signal and convenient for spotting brain haemorrhages. Diffusion MRI consists in assessing the diffusivity coefficients of water molecules in the white matter, which gives a proxy of the white fiber bundles that compose it. These bundles can then be reconstructed using tractography techniques, but they make lots of errors due to the insufficient precision of MRI to assess micro-scale structures. Computed tomography is another imaging technique which is cheaper than MRI, but it comes with drawbacks such as participant exposure to X-rays and less spatial precision in one direction (slices).

Non-living tissue can be captured from ex-vivo brain (rare material) or from biopsies, often used to diagnose tumours. Histology techniques are used to inspect such tissue at a cellular level. Tissues are sliced into thin layers, coloured using a contrast agent, visualised with a microscope and digitised by a camera. It is often used as a validation technique to verify an hypothesis derived from macro-scale observations, such as tumour grade or axons direction in a white fibre bundle for instance.

Genetic markers. There are multiple genetic markers that can be assessed. We can differentiate between structural and functional genetic, where structural will focus on capturing information about the healthy DesoxyriboNucleic Acid (DNA) of participants, and functional genetic describes how genes express in different tissues, which can vary over time and depends on the investigated tissue or cell.

Techniques to capture DNA code, which does not change during life, include Whole Genome Sequencing (WGS), which consists in reporting every nucleotide base of both DNA strands of a participant. It is however quite expensive and requires a huge storage space, as human DNA contains around 3 billion nucleotides. Moreover, it reports very redundant information across participants, because most frequent DNA variants found across individuals are known and have been documented, even if it is continuously evolving. These variants are essentially (for 90% of them) Single Nucleotide Polymorphism (SNP), where only one nucleotide base is altered at a specific DNA locus with respect to a reference sequence. There are more than 30M SNPs reported in the human genome. SNPs are well documented and there exist DNA arrays specifically designed to capture and report frequent SNPs (happening in at least 1% of the population) from a biological sample such as blood, hair or epithelial cell. Other frequently assessed genetic variations are Copy Number Variations (CNVs). A lot of the genetic material is composed of repetitions of nucleotide chains. The number of repeats can vary across individuals with duplication or deletion, which are described by these CNV data.

Functional genomic includes a wider range of measures, because there are multiple molecular processes involved in gene expression, modulation and interactions. It regroups transcriptomic

(gene expression), proteomic (protein production) and metabolomic (metabolites, i.e. larger molecules responsible for larger scale processes such as sugar or acides). Transcriptomic regroups genomic assessments for gene expression. Gene code for other biological processes such as protein production which express differently in different tissues. Transcription refers to the step where DNA is read to produce a messenger RiboNucleic Acide (mRNA), which in turn will be translated into proteins. mRNA, which is different across tissues, is a proxy of gene usage by the cells of a tissue. The quantity of mRNA can be measured using microarrays, which target specific genes. These mRNA can be sequenced as well, to better understand the gene expression. It is often used to assess gene expression in tumoural tissue obtained through biopsy, in order to identify genes involved in the cancer development. Proteomic and metabolomic require the use mass spectroscopy techniques to detect protein expression / interactions and metabolite concentration respectively.

Demographic or clinical data. This type of data includes all assessments related to the patient that is not a direct measure of its biology. Age and sex are always included in population cohorts as phenotypes, because they often have interactions with the studied pathologies. Moreover, having access to these measures allows analysts to take into account any bias or sub-represented population in the cohort and properly adapt the analysis pipeline to lessen the effect of such bias on the results. Participants medical antecedents, parent's antecedents, origins, social condition, whether they smoke and other environmental factors are often accessed through questionnaires. They allow analysts to be aware of any social or origin bias in cohorts, and study comorbidity with environmental factors. They can be very important, for instance smoking leads to more lung cancers, and an unstable socio-educational frame during childhood or taking drugs during the adolescence increase the chance of developing a psychiatric disorder such as Bipolar Disorder (BD) or schizophrenia.
Finally, phenotype data include all participant assessment related to the research question. When the study focuses on a particular disease, it should include a diagnostic (when available) as well as all the measures that led to this diagnostic, such as radiologist notes, or symptom questionnaires outcomes.

Metadata. All the data regarding acquisition protocols, such as location, chip or MRI machine manufacturer, type of machine, acquisition parameters, date of acquisition, software and any other factors that can have a potential influence on the acquired data. In particular, onset timings in task-based acquisitions, which reports when the signal was provided to the participant, when and what he/she responded, is absolutely mandatory, otherwise the data is not usable. All these metadata are very important, especially in large population imaging cohorts where data are shared and the effect of varying acquisition settings needs to be properly handled.

With all these existing assessments in mind, population studies never assess all these data, because it is expensive and not always relevant. In the following section, we describe some well known existing cohorts, precising what scientific question they address and the type of data they include.

1.2.2 . Available cohorts

We focus on openly available population imaging cohorts designed to study brain diseases or that can be used is this context. In the following, we highlight a few of these cohorts, that we identify as case / control studies opposed to general purpose cohorts.

Case / control studies. Such cohorts usually provide diagnostic for a given pathology for every participant. They are often targeting a specific pathology or disorder spectrum. We can identify cohorts designed to study oncology and others to study psychiatry.

Oncologie cohorts or dataset include BRATS [252], that was design to benchmark multimodal tumour segmentation using brain MRI from patients with low to high grade glioma. Tumours were segmented by four experts using T2-weighted images. T1-weighted, T1-weighted with enhanced contrast using Gadolinium and FLAIR were available as well with different resolutions, for 65 patients. It was then extended to a larger (2000 participants) dataset, including other assessments such as methylation status of a tumour's promoter. In fact, completion of this dataset was obtained by combining various other open oncology datasets, such as collections from The Cancer Genome Atlas (TCGA) GlioBastoma Multiforme (GBM) [308] and Low Grade Glioma (LGG) [279]. The TCGA [248, 365] is an initiative to better understand molecular processes responsible for cancer. Different types of brain tumour are included in this project, such as GBM and LGG. Biological assessment include DNA copy number, genotyping and gene expression. They were further extended to assess preoperative MRI imaging in different contrasts and postoperative histopathology.

With regards to psychiatry, cohorts do not usually include functional genomic because it is too invasive and risky. They usually include multiple contrast MRI, relevant phenotype data and sometimes genotyping. Alzheimer's Disease Neuroimgaging Initiative (ADNI) [282] was launched to find neuroimaging and chemical biomarker of abnormal ageing trajectories in old population (between 55 and 90), with more than 800 subjects, some of which with normal ageing, other with mild cognitive impairment and the rest with diagnosed AD. This longitudinal study includes different imaging techniques (MRI and PET) scanned at different visits, as well as other measurements, including cognitive abilities through various clinical tests, blood and CSF measurements. It is one of the rare large longitudinal cohort with 5 planned followup screenings over 3 years. ENIGMA [73, 72] is a consortium for grouping research efforts around the common purpose of understanding psychiatric disorders. It includes data and analyses sharing agreements. They aim at including all sorts of available imaging and genetic data. Many other and smaller initiatives exist to study specific psychiatric diseases, such as EU-AIMS [263] and ABIDE [98, 99] for autism, BSNIP [193], SchizConnect [360] for schizophrenia and BD and ADHD-200 [256] for Attention Deficit Hyperactivity Disorder (ADHD). These cohort all include various MRI contrasts, including structural MRI and functional MRI, as well as genotyping (SNP or CNV) data for some of them. In all these settings, imaging data are acquired at different sites, which tends to make difficult their analyses, as each MRI machine have different settings and acquisition conditions.

General purpose cohorts. Other cohorts exist with more general purposes, which do not necessarily include diagnoses, although they can contain unhealthy participants. Among them is the

UK Biobank [48], which regroups more than 500.000 participants from the United Kingdom. Among them, around 40.000 have already been scanned (and 100.000 should be scanned in total) in MRI machines for different imaging contrasts, including structural and functional MRI. All the participant will be assessed for a wide panel of phenotypes, including cognitive functions, environmental factors, physiological measurements and medical history, as well as genotyping including SNP data. This cohort was built to minimise the differences in acquisition protocols, using the same MRI machines with the same settings and the same genotyping chip for all participants. The Human Connectome Project[3] aims at better understanding human brain functional and anatomical connections and their relationships with different diseases such as psychiatric disorders. It includes mostly imaging data, with various functional imaging technique such as fMRI, EEG and MEG and anatomical MRI with a focus on diffusion MRI. It regroups various projects with different aims, spanning populations with varying age ranges, from infancy to old age.

All these large scale initiatives have in common to regroup images scanned at different acquisition sites or genetic data acquired using different chips. This is very convenient to gather a large quantity of data but comes with drawbacks as well.

1.2.3 . Major challenges

Population imaging present many advantages, which include regrouping a large panel of participants with multiple assessments. They provide opportunities to shed light on biological phenomena with small macro-scale effect size and increase statistical power of analyses conducted using them [240]. However, they come with different challenges. These cohorts can not be gathered at the same imaging centers, which introduces unwanted non-biological variability in images, called site effects. Additionally, although increasing the number of participant should lead to more reproducible findings, a concomitant reproducibility crisis started in 2010's [22], and the neuroimaging community was not exempt of it. We believe population imaging has the potential to change this. And lastly, translation towards clinical practice is almost never done, and comes with challenges of its own.

Harmonisation of acquisition site effects. Non-biological artefacts are inevitably introduced when gathering imaging acquired at different sites. These site effects are due to differences in scanner (manufacturer and settings), acquisition protocols and conditions, but also participant recruitment, which can have interactions with age or diagnostic for instance. As illustrated in Fig. 1.3, T1-weighted MRIs regrouped from multiple studies and preprocessed the same way display as most important sources of variations biological age and site effect. These confounding factors can be found in genomic data relying on different microarrays as well, which can be due to differences in chip, manipulations or laboratory conditions. They are refered to as batch effects. Many methods have been proposed in the literature to handle these effects, site effect harmonisation initially taking inspiration from batch effect harmonisation in genomic using bayesian modelling proposed by ComBat [121]. It has since been extended with different techniques offering other possibilities, relying on DL or conventional statistical learning [175]. It is still an open research question as there is no consensus as to which method should be used and how to validate them, depending on the

[3] Project launch note

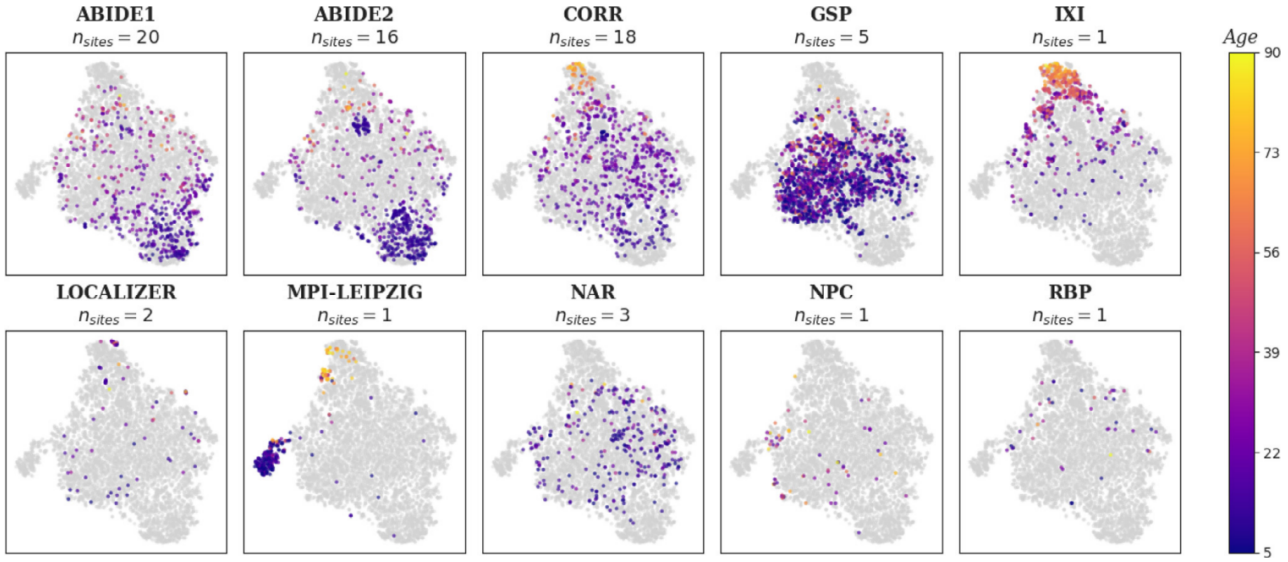


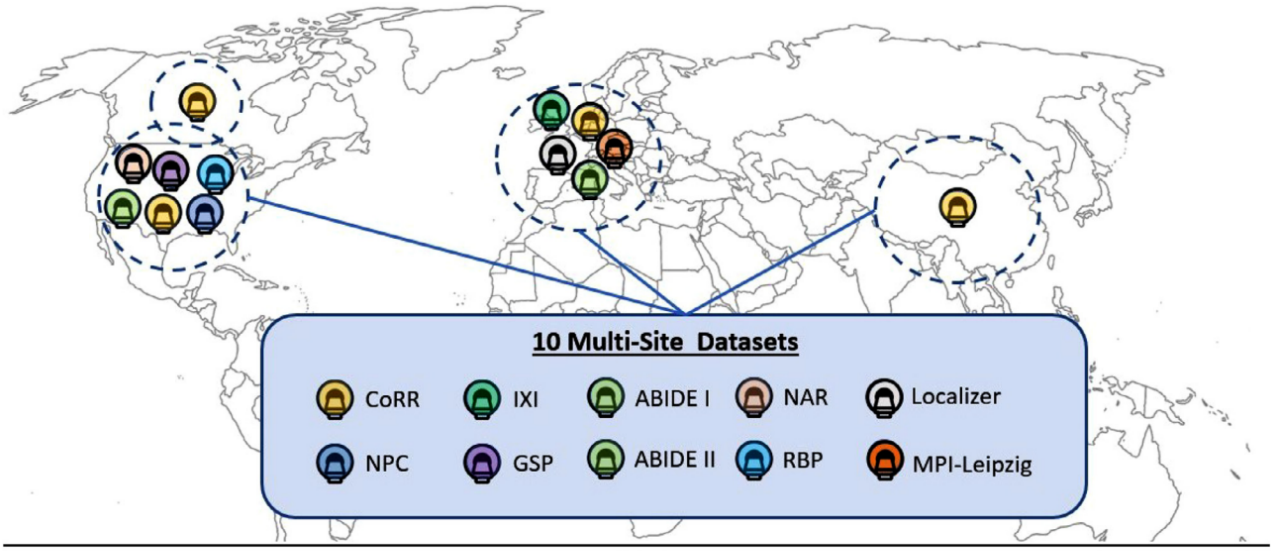
Figure 1.3: t-SNE visualisation of T1-weighted MRI from different studies pulled together and preprocessed the same way using VBM and normalised by their total intracranial volume. Image extracted from [105].

cohort settings.

In fact, there is little work able to explicitly model this site effect in images and properly remove it (or make a downstream analysis pipeline invariant to it) from a site from which no image was used to describe this specific site's impact on the produced images. This is likely due to the fact that this effect is complex, multifaceted and very specific to each acquisition site. However, we hope that regrouping a large quantity of images stemming from multiple imaging acquisition site can allow to model it. This is why we proposed the Open Brain Healthy Bank (OpenBHB) cohort, regrouping images of more than 5000 healthy participants from 10 open worldwide studies across more than 60 acquisition sites (see Fig. 1.4), along with the OpenBHB challenge [105]. This challenge, illustrated in Fig.1.5, aims at providing a shared and common benchmark for properly handling site effect with a proxy prediction task, that is learning data representations for predicting brain age from T1-weighted MRI, including different preprocessings. These representations should not be biased by site effect (i.e. bad at predicting it), while being good at predicting age on images from both seen and unseen sites. The challenge is hosted on a platform[4] on which anyone can submit a model to extract representations from the preprocessed T1-weighted MRI images, that will then be evaluated on a private test set with images from site provided in the open released dataset and others from unseen sites for age prediction.

Reproducibility. Reproducibility can refer to reproducing the results of an experiment in the exact same conditions, e.g. same data and same modelling for instance, or in different conditions, conducted by a different team, on different data or with different modelling. The most effective

[4]Here are the OpenBHB dataset website and RAMP Paris-Saclay platform to submit models.



Study	# Subjects	# Scans	Age	Sex (%M)	# Sites	# Settings	Modalities (others available)
ABIDE I	527	527	17.2 ± 7.7	82.0	20	1	T1w, (rfMRI)
ABIDE II	555	555	14.9 ± 9.5	69.5	17 (1 out)	1	T1w (rfMRI)
CoRR	1368	1368	25.9 ± 15.9	49.4	26 (8 out)	1	T1w, (rfMRI)
GSP	1570	1570	21.5 ± 2.9	42.4	5	1	T1w, (rfMRI)
IXI	558	558	48.7 ± 16.4	44.4	1	3	T1w, (T2w, DWI, PD)
Localizer	76	76	24.5 ± 6.6	42.7	2	1	T1w, (fMRI)
MPI-Leipzig	282	282	35.6 ± 17.8	60.3	1	2	T1w, (rfMRI)
NAR	289	289	22.1 ± 4.9	41.9	3	1	T1w, (T2w, fMRI)
NPC	64	64	26 ± 4.2	55	1	1	T1w, (T2w, DWI, rfMRI)
RBP	41	41	23.1 ± 5.0	48.8	1	1	T1w, (T2w, DWI, rfMRI)
Total	5330	5330	25.3 ± 15	52.1	71 (9 out)	13	T1w

Figure 1.4: OpenBHB dataset regrouping more than 5000 healthy participants regrouped from 10 open cohorts, with different population age distributions and varying number of acquisition sites. Image extracted from [105].

way to validate biomarkers is to replicate them in independent cohorts and / or different settings. You can note that reproducibility is closely related to the formerly introduced harmonisation of site effects issue, as biomarkers found without properly handling this effect are likely to not replicate in different configurations, due to the large variance explained portion attributed to site effect (see Fig. 1.3) as compared to small effect sizes of associations between brain images and phenotypes [240].

Reproducibility is a very common and generalised issue to all scientific findings. It refers to the fact that presented results in a research report can not be reproduced. There are many different types of reproducibility, and there causes are plural. Often identified causal factors could include research valorisation, which would lead researchers to voluntarily (or involuntarily) artificially augment (e.g. cherry picking or p-value hacking) their result due to pressure to publish. Other factor are data sharing difficulties and poorly designed experimental settings such as test set contamination with training examples in statistical learning pipelines.

In order to improve reproducibility of their study, authors are encouraged by journals and a few

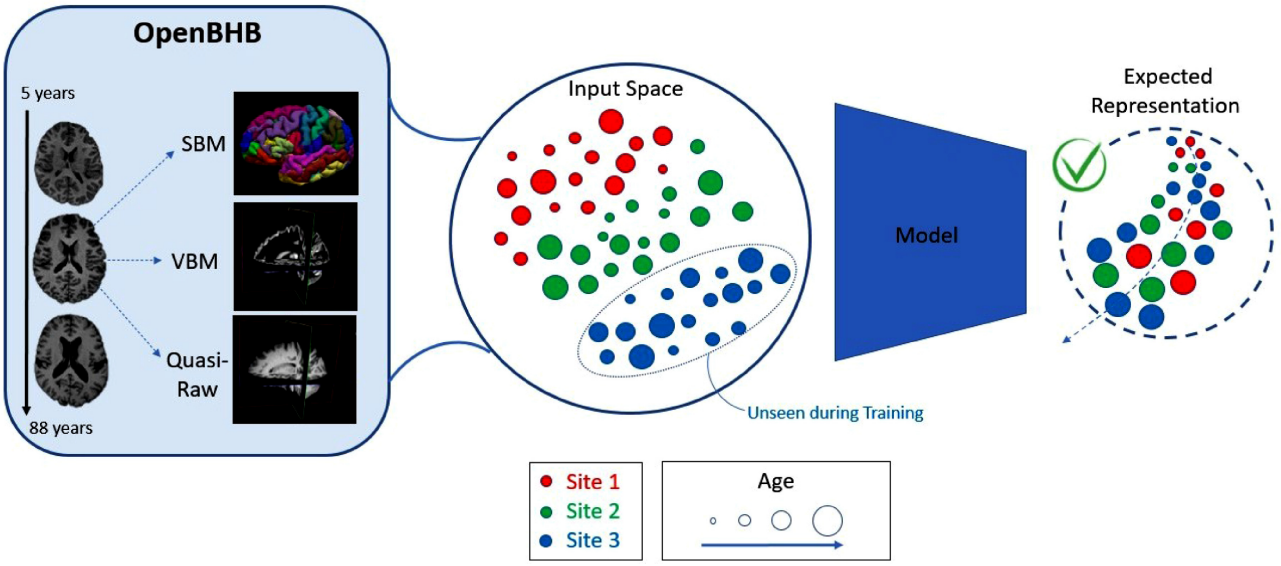


Figure 1.5: OpenBHB challenge overview. Its goal is to provide a model able to extract from various preprocessing representations good at predicting age, without being biased by site effect. Picture taken from [105].

conferences to make available their code, their data and other material that can help reproduce their results. Data statistics and preprocessing should always be provided and detailed. These initiatives happen in a wider open science point-of-view which is advocated by many to improve reproducible science.

From research to clinical routine. Another worth noting constat is that research data is very much cleaner than routine clinical data, and it has been shown that these differences can really hurt performances of current diagnostic tools, which were moslty developed using population cohorts designed for research purposes, with precise and framed data acquisition protocols. For instance, the Assistance Publique des Hopitaux de Paris (AP-HP) have been saving all their patients data in a data warehouse. A study has tried to use this T1-weighted image from this data to predict the Dementia [37]. Additionally to having numerous more difficult considerations such as building the diagnostic phenotype from the tenth revision of International Classification of Diseases (ICD-10), they showed that multiple classifying pipelines loose 15% accuracy compared to model trained and evaluated on research dataset. They highlight throughout different experiments the harmful effects of image poor quality and differences in image acquisition biasing the classifiers. This proof-of-concept paper aims at challenging the overestimated prediction rates of brain disorders obtained on research datasets and highlights the difficulties occurring when translating decision aiding tools to clinical routine data.

2 - Data-based biomarker discovery

Contents

A wide variety of tools (e.g. MRI and microarrays) are available to gain insight into the brain structure and functioning. Researchers have started to find biomarkers of the brain responsible for cognitive function or brain pathology. Some pathological markers can be found by a human expert from images: a brain tumour is often identifiable by a radiologist. Others are much more subtle, and we can not identify them by simply looking at the MRI (e.g. subtle radiomic features variation to qualify the tumour). To help medical experts to understand the underlying mechanisms of pathologies or functions, medical imaging analysts have seized data-based learning algorithm, that can be directly applied to find biomarkers.

2.1 . Machine Learning

These approaches tend to model underlying mechanisms that govern the data, in order to be able to properly link observed phenotypes with biology (e.g. Alzheimer's conversion with acceleration of grey matter shrinkage). To do so, we investigate the data X as a random variable, that follows an unknown probability distribution $\mathbb{P}(X)$. In practice, we do not know $\mathbb{P}(X)$, but we usually have access to realisations $\{X_1, \dots, X_N\}$ of X. In Machine Learning (ML), we try to approximate $\mathbb{P}(X)$ using these realisations. How we approximate this unknown distribution differs according to the question we ask and the setting we consider. In the following, we will present two very general concepts explored by ML techniques: supervised learning when annotations are available to supervise the learning algorithm, and unsupervised learning when the learning algorithm consist in describing the data density. These concepts are also illustrated by a few well-known ML algorithms.

2.1.1 . Supervised learning

In supervised learning, we usually consider i.i.d. observations $\{X_1, \dots, X_N\} = \{(x_1, y_1), \dots, (x_N, y_N)\} \subset \mathcal{X} \times \mathcal{Y}$. We want to model their joint probability $\mathbb{P}(X, Y) = \mathbb{P}(Y|X)\mathbb{P}(X)$, following the Bayes rule. The question we address in supervised learning is whether or not we can predict y when observing x, more formally if there exist a function f such that $f(x) = y$. Let $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$ be a loss function, evaluating how far predictions $f(x)$ are from y. We try to find a function f such that it minimises the expected theoretical risk under the joint distribution $\mathbb{P}(X, Y)$.

$$f = \text{argmin } \mathbb{E}_{\mathbb{P}(X,Y)}[L(f(X), Y)] \tag{2.1}$$

We usually relax this untractable risk by searching for an estimator f that solves the Empirical Risk Minisation (ERM) problem $R(f) = \frac{1}{N}\sum_{i=1}^{N} L(f(x_i), y_i)$. For some restricted classes of estimators f with specific loss function, the solution to this optimisation problem is analytically tractable, e.g. for linear functions with the squared error as loss function. However, this is in general not the case. Another way of finding a minima of the ERM is by using a gradient descent when L is convex.

We refer to classification when $\mathcal{Y}$ is finite, i.e. the variable to predict $y \in \{0, \dots, c\}$, c being the number of classes, and regression when $y \in \mathbb{R}$. A parametric class of estimators is defined as $\{f_\theta, \theta \in \Theta\}$, with $\Theta \subseteq \mathbb{R}^k$, $k \in \mathbb{N}^*$. Below we introduce two widely used supervised machine learning models.

K-Nearest neighbours. The K-Nearest neighbours (K-NN) is a non-parametric estimator that predicts labels y corresponding to observation x according to the labels from the K nearest neighbours of x. More formally, $f(x) = g(\{y_i, x_i \in N_K(x)\})$, where $g : \mathcal{Y}^K \to \mathcal{Y}$ is a function to aggregate the labels from the K neighbours $N_K(x)$ of x. It requires $\mathcal{X}$ to be equipped with a distance function, e.g. the Euclidean distance in $\mathbb{R}^p$, to define neighbourhoods. For instance in a regression settings, the function g can be defined as $g((y_1, \ldots, y_K)) = \frac{1}{K}\sum_{i=1}^{K} y_i$, where the predicted value $f(x)$ is the average value of the K nearest neighbours of x. In classification settings, g could be the majority voting class.

Generalised linear model. The Generalised Linear Model (GLM) is the simplest model that can be used as a prediction model. It relies on the assumption that $y \sim \mathbb{P}(Y|X)$ is an exponential distribution. The estimator function f, which is the mean of this distribution, is expressed as $f(x) = g^{-1}(x\beta)$, where $g^{-1} : \mathcal{X} \to \mathcal{Y}$ is an invertible link function. In regression settings, $\mathbb{P}(Y|X)$ is assumed to be the Gaussian $\mathcal{N}(f(x), \sigma^2)$, and g is simply the identity function. For classification, the most widely used GLM is the logistic regression, which assumes $\mathbb{P}(Y|X)$ is a categorical distribution function, and uses the logit as link function, i.e. $g^{-1}(x) = \frac{1}{1+e^{-x}}$.

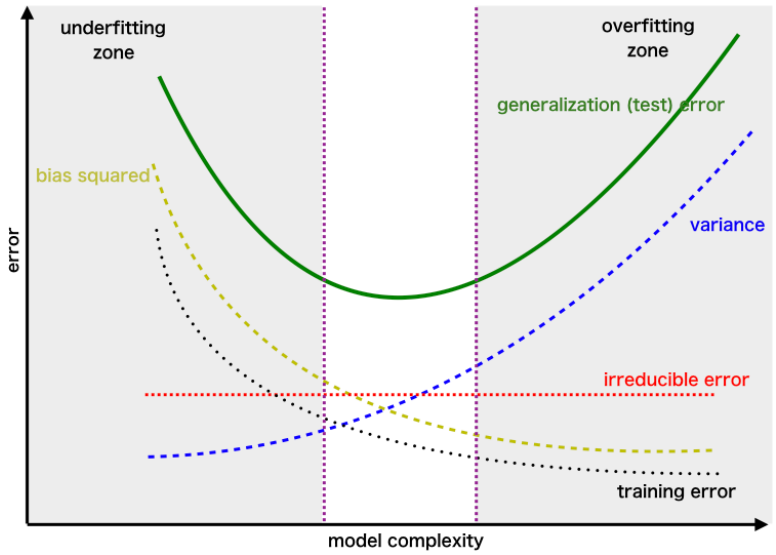


Figure 2.1: Illustration of the bias-variance tradeoff. Picture taken from an external source[1].

Regularisation. In machine learning, we hypothesise the estimator's form by approximating it within a class of functions. The final objective of this learning algorithm is to perform well on unseen examples, which is characterised by the generalisation error, i.e. the value of the empirical risk on unseen examples. This generalisation (or test) error can be decomposed into two components, namely the bias-variance decomposition (illustrated in Fig 2.1). The bias refers to the part of the error due to the difference between the predicted values and their true values due to the hypothesised

[1]Link to the article.

class of function, and the variance is the error due to the sensibility of the estimator to small variation in the input. There is often a trade-off between bias and variance, which can be set by monitoring train and test (or validation) error during training. While train and test errors diminish, we usually are in an underfitting senario, whereas when test error increases while train error continues to decrease, we call this overfitting. A good way to reduce overfitting on the training set, i.e. reducing the variance part in the generalisation error, is to penalise the estimator during training, in order to restrict the class of functions it can approximate to smoother functions, such that it can be less influenced by small variations in the data. We introduce below some of them, usually applied to parametric estimators f_θ by simply adding the terms to the loss function optimised to fit the model. Penalty terms are usually weighted using a hyperparameter λ, which sets the strength of the regularisation during training.

Ridge. A commonly used regularisation is the l_2 or Ridge penalisation term, that is formulated for parametric models as $\|\theta\|_2^2$, where θ are the real-valued parameters of the estimator. This penalty term ensures that the weights of the model are constrained to stay close to zero, penalising more high than small amplitudes.

Lasso. Another standard penalty is called Lasso [346] or l_1 penalisation. For parametric models, it can be written as $\|\theta\|_1$, θ being the real-valued parameters of the estimator. This term forces all the parameters to stay equally close to zero, producing a selection mechanism: terms that are unnecessary for the model to produce good predictions will go to zero. The associated λ configuring the strength of the penalty can be seen as a sparsity parameter.

2.1.2 . Unsupervised learning

In unsupervised learning, there is no label Y, we only have access to observations $\{x_1, \dots, x_N\} \subset \mathcal{X} \subset \mathbb{R}^p$. These learning algorithms can be applied to a wider range of problems than supervised learning, because it does not require annotations. In fact, the learning paradigm here will try to approximate the distribution $\mathbb{P}(X)$. Let's assume f^* is the density function of $\mathbb{P}(X)$. Considering the Kullback-Leibler divergence between two distributions q et g $D_{\mathsf{KL}}(q\|g) = \int q(x) \log \frac{q(x)}{g(x)} dx$ as the expected risk, the Bayes rule says that:

$$f^* \in \operatorname*{argmin}_g D_{\mathsf{KL}}(f^*\|g)$$

We have $D_{\mathsf{KL}}(f^*\|g) = \int f^*(x) \log f^*(x) dx - \int f^*(x) \log g(x) dx$, and since the first does not depend on g, we have:

$$\begin{aligned} f^* &\in \operatorname*{argmax}_g \int f^*(x) \log g(x) dx \\ &\in \operatorname*{argmax}_g \mathbb{E}_{\mathbb{P}(X)}[\log g(X)] \end{aligned} \tag{2.2}$$

This comes back to the empirical risk minimisation problem highlighted in Section 2.1.1, considering the loss function defined as $L(f, x) = -\log f(x)$. We call a maximum likelihood estimator $f \in \mathcal{F}$ a

probability density function such that:

$$f \in \underset{g \in \mathcal{F}}{\text{argmax}} \frac{1}{N} \sum_{i=1}^{N} \log g(x_i)$$

There are many different approaches to solve this issue, each of them relying on different hypotheses. We introduce a few of them in the following.

Dimensionality reduction

This type of unsupervised methods hypothesise that there exist a lower dimensional manifold within the $\mathcal{X}$-space which concentrate a high density of X. Such techniques can be linear such as Principal Component Analysis (PCA) or non-linear such as t-SNE [237]. We introduce a few widely used ones.

Principal components analysis. The PCA relies on the singular value decomposition of the centred observations $X = [x_1, \dots, x_N]^\mathsf{T} \in \mathbb{R}^{N \times p}$. In particular, their exist orthogonal $p \times p$ matrices U and V such that $X = UDV^\mathsf{T}$, where D is a diagonal matrix with diagonal elements $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ being the singular values of X. The first few columns of UD generate a lower dimensional vector space. Each of these columns represent principal components, that express a decreasing variance of X, starting from the first column which describes the linear projection dimension maximally expressing data variance.

Multidimensional scaling. Rather than using the observations themselves, MultiDimensional Scaling (MDS) solely uses their dissimilarity. It tries to find a lower-dimensional manifold in which dissimilarity between observations are maximally preserved.

There are many dimensionality reduction techniques. They are nowaday often used as visualisation tools, in order to display in a low-dimension space (usually 2D or 3D) captured semantic in higher dimensional data representations. Among them, PCA, t-SNE and UMAP [247] are the most often used.

Clustering

Such algorithms try to find some convex regions of $\mathcal{X}$ that contain modes of $\mathbb{P}(X)$ [156]. They aim at creating somewhat homogeneous groups (or clusters) of observations, hypothesising $\mathbb{P}(X)$ is a mixture of simpler densities representing these different groups. In the following, we introduce two of the most iconic clustering methods, namely the K-Means and the Gaussian Mixture clusterings.

K-Means. This algorithm does not explicitly formulates a distribution $\mathbb{P}(X)$. Rather, it tries to find clusters using central points in an iterative process, with a fairly simple heuristic: the average dissimilarity between observations and their centre point within clusters should be minimal. It uses the Euclidean distance to evaluate dissimilarity between observations. Initial random cluster assignment influences the outcome.

Gaussian Mixtures. Such clustering algorithms explicitly hypothesise that $\mathbb{P}(X)$ is a mixture distribution, i.e. $\mathbb{P}(X) = \sum_{k=1}^{K} \pi_k \mathcal{N}(X|\mu_k, \Sigma_k)$, implicitly introducing a K-dimensional binary latent variable $Z = (Z_1, \ldots, Z_K) \in \{0,1\}^K$ such that $\sum_k z_k = 1$ with $z = (z_1, \ldots, z_k)$ realisations of Z, and that $\mathbb{P}(Z_k = 1) = \pi_k$ with $\sum_k \pi_k = 1$. This formulation gives a convenient form to the likelihood function that has explicit partial derivatives with respect to the parameters of the models. The maximum likelihood estimator is usually computed using the Expectation-Maximisation (EM) algorithm to fit its parameters [35].

We only scratched the surface by introducing a few key concepts of machine learning and some of its most known algorithms. We could have discussed decision trees or kernel based methods as well.
A lot of classical machine learning algorithms have the convenient property of being relatively interpretable: linear models (such as canonical correlation analysis or logistic regression) can be inspected using their weights, principal component analysis looking at directions with high variance, random forests inspecting decisions, and so on. There even exist mechanisms to select most relevant features to solve the learning task such as the l_1 regularisation.

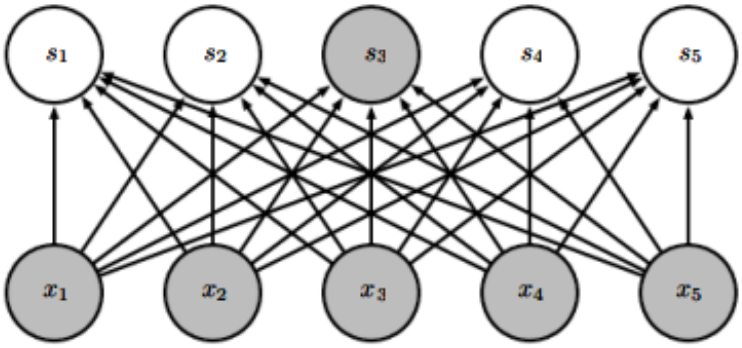
However, a paradigm shift has been observed in the 2010's. Artificial NNs had been there for quite a while now [302], but only then have researchers begun to propose architectures capable of surpassing all other machine learning models on specific tasks, in particular in computer vision using Convolution Neural Networks (CNNs) [68, 207, 321]. This opened up a whole new chapter in machine learning: NN based learning, also called Deep Learning. Nowadays, it is widely deployed in almost every domain where machine learning is used. In particular, it offers very flexible learning algorithms and shows good abilities to exploit data structure to automatically learn relevant features without the need of explicit expert feature engineering.

2.2 . Deep learning

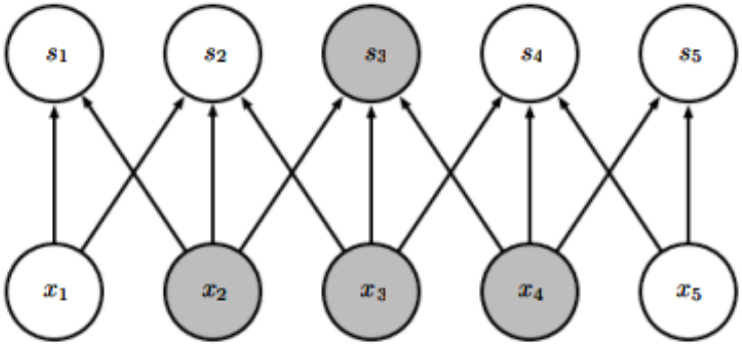
This sub-field of machine learning regroups all the learning algorithms using ANNs. They were introduced in the early 60's with Rosenblatt's MultiLayer Perceptron (MLP) [302]. A MLP f is a succession of linear projections, called dense layers, interspersed by non-linear activation functions. You can find a schema of a dense layer in Figure 2.2. For instance, $f(x) = f_1 \circ f_2 \circ \ldots f_K(x)$ is a K-layer MLP, with $f_k(x) = \phi_k(\beta x + b)$ where ϕ_k are activation functions. Today, ANNs are trained using Stochastic Gradient Descent (SGD) [293, 10] or their extensions [195] through backpropagation [227, 366]. Their use and application was slowed down by its high requirements in hardware and data. Computing on Graphical Processing Units (GPU) demonstrates great performances for matrix operations, and NN implementations were adapted to be trainable on such hardware [272, 58]. The advent of large open databases such as ImageNet [92] allowed prediction performance increase in a large panel of tasks, by providing a shared resource to collaboratively develop NN architectures and assess their performances.

However, due to their intractable complexity, NN theory is poorly understood. A very surpris-

ing phenomenon is their astonishing generalisation properties, which scales with their complexity, unlike other ML model, which challenges the until then validated and illustrated bias-variance decomposition in Fig. 2.1 to explain ML generalisation as a function of model complexity. Therefore, NN expressivity properties have been described further to understanding their generalisation properties. It seems that their data adapted operation such as convolution operators implement inductive biases which could be one source to explain these generalisation properties. Other than their surprising generalisation characteristics, interpreting NNs is not straightforward in practice. We try to describe a few explanation methods that tries to overcome this issue. And finally, we introduce an unsupervised concept that has gained a growing interest with NNs called self-supervised learning. It is grounded in learning representations enforcing inductive biases to the NNs by exploiting carefully designed data manipulations.



(a) Dense layer



(b) Convolution layer

Figure 2.2: Schematic illustrations of NN layers. Black lines represent the parameters of the layers. (a) Dense (or fully-connected) layers connect all inputs to all output features (matrix multiplication). This is the type of layer that compose a MLP. (b) Convolution layers only connect neighbouring features. The represented layer is a 1D convolution kernel of width 3. We highlight the receptive field of the output feature s_3 in grey, meaning the inputs from which it is issued. Credits to [138].

2.2.1 . Expressivity of NNs

In the late 80's, universal approximation theorems [82, 173] show that a 2-layer MLP can approximate any continuous function on a bounded space. It has been extended since to other types of neural networks and generalised to any function. Neural networks are parametric universal estimators, but we do not yet know why they generalise so well [385]. This shows that machine learning theoreticians do not have the proper tools to understand the great generalisation properties of NNs, which is still on open research question. In particular, people argue that such generalisation properties would come from an implicit regularisation inherent to SGD [155], called Simplicity Bias [18, 349], encouraging NNs to approximate the simplest function to solve a task. It was shown that this concept is not enough to explain generalisation of neural networks, and can in fact hurt their performances when NN overfits on simple features and can no longer detect features of interest [315]. A way to improve generalisation of a NN architecture has consistently been to improve the inductive biases it implements [144]. We refer to inductive biases every aspect of the learning algorithm that has an effect on the shape of the approximated estimator f, which are the properties of the explored functions space (i.e. architecture of the NN), the loss function and the learning algorithm (e.g. SGD).

2.2.2 . Structure modelling and data-adapted operators

In this section, we focus on discussing how NNs model data structure by exploiting inductive biases for different data types. Considering geometrical deformation D of an input x, inductive bias can take the form of invariance to such deformation. The predictor function f is invariant to D when for all $x \in \mathcal{X}$, we have $f(D(x)) = f(x)$. This means that the output of the function f is not impacted by the deformation D of the input. This is a good property for a classifying network on natural imaging, mapping image to one class among others. Usually the label's image should not be altered via translation for instance. However, considering the problem of image segmentation, translation invariance is not a desirable property, because you would want the predicted class of the pixels to shift with the translated image as well. This property is called equivariance. A function f is called equivariant to a geometrical deformation D when for all $x \in \mathcal{X}$, we have $f(D(x)) = D(f(x))$. We introduce below some of these inductive biases and associated operators, along with the type of data they are usually used to model.

Convolution. Convolutions were first introduced in 1980 [123] as an operator able to mimic the hierarchy of cells in the visual nervous system [178], able to recognise patterns based geometrical resemblance and not affected by shifts. Convolution layers are composed of kernels. For simplicity, let's assume we are dealing with 2D image data. A convolution kernel is a rectangular patch with learnable weights. It can have a third dimension if the image data has multiple channels (e.g. RGB images). When applied, this patch slides over the image and outputs the dot product between the patches weights and corresponding image pixels. This mechanism allow to the kernels to learn features maps equivariant to translation. Moreover, this strategy reduces drastically the number of weights of the model, as illustrated in Figure 2.2, and parameters are shared across features. Convolution layers are often followed by a pooling layer. Its role is to reduce the spatial dimension of output feature maps from the convolution layer. It usually considers small areas of the output feature maps and aggregates them into unique values, using a mean or a max function.
CNNs are translation equivariant because the convolution operator is translation equivariant by def-

inition. CNNs can display translation invariance properties as well, inherited from the pooling operations. This invariance is not as interesting as equivariance, because it reflects the fact that these networks erase some information, notably the position of identified patterns. CNNs were first applied to images for classification tasks, where such properties are very interesting, and this is were convolutions have been applied the most up to today. You can find an illustration of a CNN in Figure 2.3. Many highly performing architectures rely on this operator [219, 321, 207, 159, 176]. They have been applied to speech processing as well, using 1D convolution [199]. Lately, they have often been use for image generation in generative models such as Variational AutoEncoders (VAEs) [197], Generative Adversarial Networks (GANs) [139] or diffusion models [168, 295, 289]. Moreover, multiple extensions of this operator have been proposed to be applicable in geometrical deep learning [40], to learn from general graph structures. The equivariance / invariance properties of these extended operators change according to their formulation and application. We discuss such operators in the Section 3.2.1. Finally, they have been used extensively for segmentation task, especially in the medical domain area, with the well-known and still widely used U-net [300] architecture. They are however not equivariant to rotation, which can be very usefull in Segmentation tasks. For instance, capsule networks [167, 306] implement equivariance to rotation or other local deformations and can show improved segmentation performances compared to convolution [212, 283].

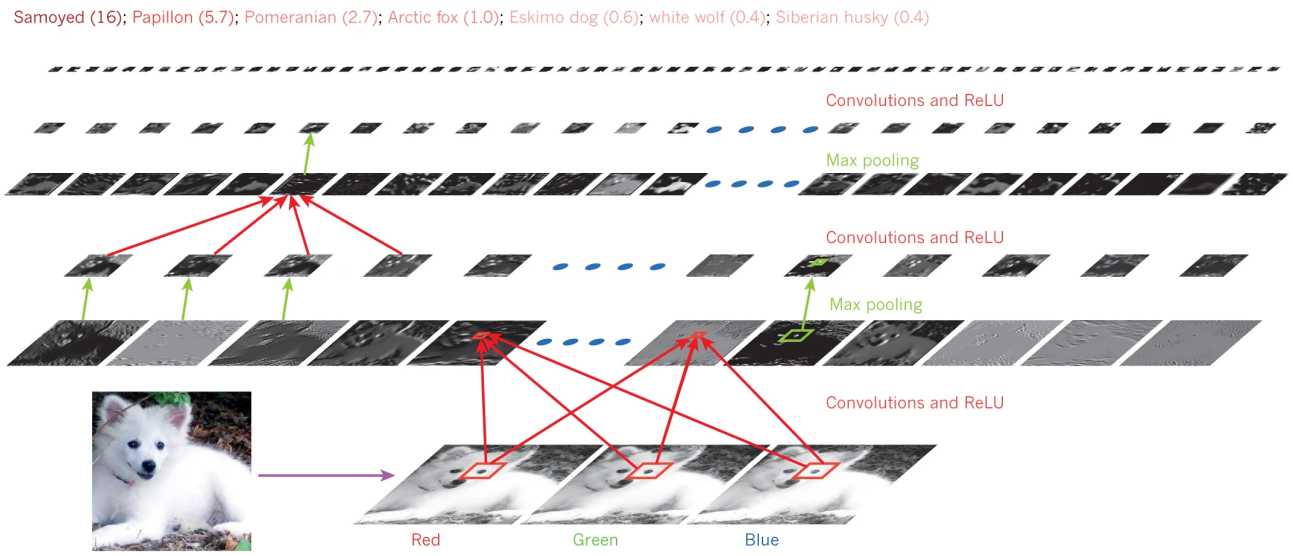


Figure 2.3: Inside a CNN. Bottom-up: input is a 2D RGB image of a white Samoyed. Convolution filters are applied channel-wise, then summed into the first layer output feature maps. The non-linear activation function Rectified Linear Unit (ReLU) is applied to these feature maps, which are then spatially reduced using a max pooling operation, and input to the second convolution layer, and so on. From the last layer output feature maps, or representations, are predicted the weights corresponding to each class. The network mostly predicts the image is that of a Samoyed. This image was drawn from [220].

Recurrent units. The inductive bias implemented by recurrent units is that an object's characteristics depends on the preceding objects within a sequence. For instance, a word's meaning depends on the previous words in a sentence. Recurrent Neural Networks (RNNs) derive from the Ising model introduced in the 20's [43]. They were then extended to be adaptable [9] and trainable using gradient descent [172, 305]. Extensions popularised these networks modelling long term dependencies using gating mechanisms [169, 67]. Such networks work particularly well for speech processing and have demonstrated better performance [145, 117] compared to conventional machine learning and became State-of-the-Art (SotA). They have been used for text related task as well, such as language translation [334], language modelling [189] and caption generation [354, 200] (image to text).

Attention. In attention mechanism, the inductive bias is less restrictive. It models the fact that an object within a group relates strongly to a few other objects of the group. Its contemporary implementation was proposed in 2017 [350] as part of a Transformer architecture for written language translation. These Transformer based architectures quickly dominated all applications related to text modelling by implementing the most famous models used nowadays (e.g. BERT [97], GPT [41], Gemini [341], Mixtral [182]). Lately they have been used in computer vision as well [102], and have shown SotA performances. Indeed, their attention mechanism allows a better use of the long range dependencies within an image, but requires a larger amount of data than their convolutional counterparts. For that reason they are not so practically applicable to domain with scarce data such as clinical medical imaging.

2.2.3 . Lack of interpretability : black-box algorithms

Neural networks are highly expressive models, able to model data structure using various inductive bias from prior knowledge. This expressivity, however, sacrifices interpretability, leading to their designation as black-box models. In linear models, we can easily track features responsible for predicting such class, i.e. question the model by directly inspecting it. This is not the case for NNs, which learn non-linear functions thanks to their activation functions. This fundamental difference led to most criticisms that neural network based models are facing.

Deep learning literature is mostly empirical: design choices are made, then tested implementing a model and testing it on a task of interest. There are only a few works on neural network theory, some of them were mentioned in Section 2.2.1, but DL research is mostly application-centred. For instance, we do not yet understand why they generalise so well, without explicit regularisation [385], as compared to other machine learning models. In addition, there is no guarantee in general that the function learned using gradient descent for the empirical risk minimisation is a global minimum. In fact, it is most often not the case, and sometimes a model is stuck in local minima "bassins" and can not escape, notably in transfer learning [265]. Most theoretical questions around deep learning have not yet been solved. Even their biologically inspired initial formulation [302] is empirical.

Additionally, some elements have brought worries to the safety of using such models. Indeed, it has been shown that is is fairly easy to trick a model into doing a wrong prediction [42], by only slightly modifying an input. This is called an adversarial attack and it has not been solved yet. Other issues related to fairness of the trained NNs has brought attention as to the bias learned by a model. Models are biased, because they can never see all the possible existing examples and outcomes.

These issues in fairness [56] have exploded when Amazon was using an automatic system based on a DL algorithm to select candidates for their recruitment processes. In particular, this algorithm was more prone to select men than women[2], because it was trained with more men profiles accepted than women, which was due to the inherent societal bias that there is more men working is such domain. The same append for COMPAS, a tool to aid decision in judiciary cases, which was shown to wrongly attribute higher risk scores to black people than white people [245]. As human are most often not aware of their biases, how can we make sure that the models we train are not ? This is still an open research question.
Of particular interest, one can reasonably wonder why a model makes a decision / prediction. For a system to aid decision designed for clinicians, the doctor usually must be aware of the markers that lead a system to make a decision. This particular issue can be referred to as interpretability of NNs.

There is no consensus or mathematical definition of interpretability [228]. One popular definition is "the ability to explain or to present in understandable terms to a human" [101]. It is often used interchangeably with *explainability*, even if some studies have attempted to clarify the concepts [228, 101]. Interpretation can come in very various forms, with very different objectives and outputs. Each of them come with specific properties and scope. Some ML models comes with built-in interpretability, that can be presented as feature contributions (e.g. linear models) or rule-based decisions (e.g. decision trees). In the following, we mostly consider interpretability methods for NNs, which are intrinsically not interpretable, due to the complexity of the functions they learn. There are many existing taxonomies [226, 51, 389] of interpretation methods. We will present below a few model-agnostic methods, i.e. that are able to provide interpretation regardless of the class of function that is being approximated by the ML algorithm [226, 51], as opposed to model-specific interpretation methods. These methods are applies post-hoc, meaning after training the model. There exist methods acting during training for improved interpretability [389], but we will not discuss them here.

Model-agnostic methods. Model-agnostic interpretability are techniques to produce explanations of any ML model. They are therefore post-hoc, since there is no assumptions on the model used. In particular, we present in this paragraph two of them, to give an intuition on the type of techniques used by model-agnostic explanation tools. Local Interpretable Model-agnostic Explanations (LIME) [292] focuses on giving interpretations of prediction score produced by any classifier. It uses a sampling scheme which generates random data in the neighbourhood of a reference instance, for which the model's prediction is available. These newly generated data are classified using the same model, and these prediction weighted by their initial distance to the reference. Then a simpler interpretable by nature model, such as a decision tree, is trained on this new dataset. By interpreting this model, we can understand locally the inital black-box classifier. Another worth mentionning method is SHAP, which stands for SHapley Additive exPlanation [234]. This game-theory inspired work proposes a unified framework for attributing feature importance with respect to predictions. It encompasses other methods such as LIME within the framework and subsequently proposes SHAP values as a unified measure of feature importance. Even though these model-agnostic methods are theoretically applicable to any ML model, they are limited to the relationship between input and output and do

[2]You can read more in this article.

not leverage the intrinsic properties of the explained models. Some NN-specific interpretation techniques try to overcome this drawback.

NN-specific methods. A very basic yet intuitive way to understand what a network learn is to inspect its weights. In MLPs, this would not lead to very interpretable results, but CNNs have filters which capture different geometrical pattern in an image for instance. We can easily visualise such patterns. Moreover, attention-based networks such as transformers [350] or vision transformers [102] can be interpreted by investigating attention maps of instances, which can provide reasonable interpretations [49]. In a more general setting, interpreting NNs has mostly been developed for CNNs in computer vision, and earliest works include gradient based feature attribution maps, also called saliency maps [320]. Later, grad-CAM [313] was proposed, using the class-specific gradient flow in CNNs, providing a coarse localisation map of important regions in the image for the classification process. Many extensions of such methods exist and they have been extensively studied. However, some works have shown that these gradient-based interpretability methods were unable to account for internal mechanisms of CNNs, as they moslty focus on the first few layers to provide interpretability [4], so they do not allow us to understand why deeper is better. This work [4] also shows that most existing methods do not reflect feature importance in the decision process, but rather a kind of edge detector. There is no consensus as to which method is the best for interpreting NNs, and it is still an active research field.

2.2.4 . Self-supervised learning

Self-Supervised Learning is a statistical learning scheme which can be classified as an unsupervised learning paradigm, since there is no requirement for human annotations in its formulations. SSL has concentrated a huge research effort in the recent years. This is particularly the case, since large databases have become openly available, and the fact that acquiring labels for every specific task we try to solve using machine learning models is very expensive. The latest and most efficient SSL techniques all use a supervisory signal called pretext task provided by the data itself, without relying on the explicit use of human labels. These pretext tasks can be defined in various ways, but we can identify commonalities and discrepancies. We distinguish in particular two types of method: generative models and joint embedding architectures [319].

Generative models. Generative models work at the input-level, and attempt to generate or reconstruct data example with a distribution close to $\mathbb{P}(X)$. Their loss usually applies in the X-space and assess how close the distribution of generated samples is to $\mathbb{P}(X)$. Such models can have different forms. The simplest ones are Auto-Encoders (AEs) and their extensions. AEs were first introduced a few years ago [166, 205], following the assumption that the data can be generated from a lower dimensional manifold. They have been extensively studied and there exist many deriving methods such as denoising AEs [352, 353]. These methods rely on the minimisation of the empirical risk, with as loss function the squared error for continuous data and cross entropy for finite X-space. A variational extension exists [197] and uses the Evidence Lower Bound as loss function, considering the empirical risk minimisation problem as a likelihood maximisation problem (see Section 2.1.2). There are many other works exploiting reconstruction as pretext task after artificially degrading inputs. Some

use noise [352], others colourisation [387] or inpainting [278, 157]. They have been characterised as Masked Image Modelling (MIM) methods [26], extending the concept of masked language modelling [97] from natural language processing to images. Other SSL methods are specifically dedicated to data generation, i.e. models learning to sample from $\mathbb{P}(X)$, such as GANs [139] or diffusion models [168, 295].

Joint embedding architectures. These SSL methods process multiple views (possibly altered versions) of an input signal through encoding networks, producing representations for each of the views. Their loss is usually applied in a low dimensional embedding space. Numerous methods have been developed, among which we can distinguish three families. Deep metric or contrastive learning [24] techniques usually rely on encoders sharing the same weights, also called Siamese networks [39], which were first applied to face recognition by learning a similarity metric between faces [65]. More recent contrastive learning techniques have shown great perfomances in SSL [60, 158], by bringing close together in embedding space views coming from the same image, while repelling them from others (referred to as negative samples). This contrastive technique is described in more details in Section 3.3.1. Other works do not consider negative sampling, because of its dependence on large batch size and hence heavy memory requirements. Rather, these methods solely focus on bringing close together in embedding space different views from the same signal. These methods have been shown to be prone to dimensional collapse [184], which happens when the embeddings span a low dimensional space (or in the extreme case, maps all the views to the same constant). Two main approaches exist to prevent it from happening. Self-distillation methods [24] use two encoders, where only one of them is updated through gradient descent, and the second encoder uses a running average of the other encoder's weights [147, 49]. Another family of methods stems from the canonical correlation analysis theory [24]. These methods [383, 28] intend to regularise the co-variance matrix of the embeddings spanned from two views of the same signal, using different criteria.

SSL methods are usually presented as unsupervised representation learning algorithms, yet they are always evaluated using human annotations on unseen data, either directly using a linear or K-NN predictor, or by fine tuning the weights of the encoder for the specific task at hand. Only a few works [134, 130] have proposed unsupervised metrics for SSL hyperparameter tuning and evaluation. In this context, the main goal of SSL algorithms is to provide good NN initialisation for further downstream tasks of interest. This is called Transfer Learning (TL) and has great perspectives, particularly in the medical field where it is impossible to have a large annotated dataset for every existing disease. Note that it was shown suboptimal to use a supervised predictive task on natural images to initialise a NN for a downstream medical task [286], due to the large domain gap between natural and medical images. An appealing alternative is to pretrain using SSL on a large cohort of healthy individuals, then transfer to a smaller clinical dataset to solve a task of interest [103].

Most of these SSL techniques rely on data augmentation. Recent joint embedding architectures all identify data augmentation as a key component to achieve good performances [60, 147, 49], and pretext tasks design in MIM [26, 24] are data augmentations as well. It acts as a mechanism to force

the networks to only learn relevant features from the data, ignoring (or being invariant to) thoughtfully designed data manipulations. They appear as yet another way to implement inductive bias in NNs. This is why data augmentation is really at the core of SSL [24], while being always domain-specific, which is one of the reasons why it is so difficult to apply it to other domains than natural image or text (other being the existence of large openly available benchmarks such as ImageNet [92]).

Part II

Contributions

We present in the following the main contributions of this thesis.
In Chapter 3, we describe how to properly integrate cortical surface measurements in NNs to properly account for their underlying biological structure, using their spherical representations. Methodological contributions and experiments are detailed.
In Chapter 4, we introduce the paradigm shift in psychiatry, which advocates for a more comprehensive approach including multiple assessments. In this context, we propose to use Deep multi-view Probabilistic Graphical Model (PGM)s to integrate these data in population cohorts. We propose an interpretability model and stabiling procedure to alleviate the different variabilities encountered with a statistical model when analysing population cohorts. This chapter focuses on introducing these different methological concepts.
In Chapter 5, we apply the proposed integration methods to the transdiagnostic cohort HBN in order to discover shared biological neural bases across psychiatric disorders. In particular, we apply the methods presented in Chapter 4 to discover transdiagnostic brain-behaviour associations.

3 - Cortical surface structure modelling

Contents

As introduced in Section 1.1.3, cerebral cortex is of major interest to study brain pathologies. The cerebral cortex has a very complex structure, that varies a lot across individuals, and the links between particular shapes of the cortex and some diseases are still incompletely described and even less understood. In fact, the brain is composed of two hemispheres, linked by the corpus callosum, where each hemisphere appears to be wrapped of uninterrupted grey matter, composing the cortex. This cortex has a highly folded structure, which allows a larger cortex surface area, developed during the last trimester of pregnancy in parallel with the final migration of neurons in the cortical plate [288, 287, 332]. Beyond a large-scale organisation of primary sulci that is consistent between individuals, albeit variable in details, it is recognised that cortical structural features are often associated with various brain pathologies [129]. With the advent of non-invasive neuroimaging, this cortical structure can be very precisely inspected using T1-weighted sMRI, and there exist tools to extract such structure from the 3D volumes, to model the cortical surface as a graph.

3.1 . The cortical structure can be described by a graph

In the following, we introduce how the brain cortex can be represented as a graph. We start by describing the cortex and some of its known properties, then we decline the mathematical notion of graph, and finally we describe how the cortex can be modelled as a graph and the existing software to extract a graph representation of the cortical surface from brain sMRI.

3.1.1 . Brain cortical surface

The cerebral cortex plays of key role in sensorimotor, cognitive socio-emotional, and mentalising functions. These unique capabilities of humans that are embedded in their cortex, derives from their evolution process [339]. Through late in-utero development, childhood and adolescence, and young adulthood, a sequence of brain developments ensures continuous progression from the initial cortical plate to the cortex organised in forms and functions. Genes and molecular mechanisms which underlie the developmental process govern the final cortical organisation, and their inter-individual variants explain a part of the mature cortex variability [304]. The interactions between the environment and the maturation of the brain are expected to favour and influence the cortex organisation [75]. It is only natural to assume it may produce abnormal forms and functions too.

In vivo MRI enables to quantify macro scale properties of cortical structure, such as cortical thickness, surface area, cortical volume. It also provides the opportunity to measure the sulcation of the cortex either by measuring various indices of sulcation or explicitely modelling the sulci shapes from the cortical surface. All these measures are proxies of the underlying biological mechanisms happening in the brain. Since technological tools have allowed us to access such biological markers of the cortical surface, people have tried to find correlates between this biological information and various psychopathological traits observed in autism spectrum disorders, depression or schizophrenia.

3.1.2 . Graphs: a mathematical notion to describe data structure

Let $G = (V, E)$ be a graph. $V \subset \mathbb{R}^p$ is the set of vertices of the graph. Each vertex has p features. $E \subset \{1, \dots, |V|\} \times \{1, \dots, |V|\}$ describe the edges between the graph vertices. Edges can be directed or undirected, hence composing directed or undirected graphs respectively. Edges can have description features as well, such as weight, but we restrain the definition to binary edges.

Another way to define a graph is to use its adjacency matrix. Let $A \in \mathbb{R}^{|V| \times |V|}$ be the adjacency matrix of G. Then we have $A_{ij} = \mathbb{1}_{(i,j) \in E}$. In an undirected graph, A is symmetrical. Let restrict ourselves to undirected graph, as the following applications and operators mainly use undirected structures.
We define the degree d_i of a node i as its number of neighbours, $i.e.\ d_i = |\{(j,k) \in E,\ j = i\}| = \sum_j A_{ij}$. Let D be the degree matrix of G. D is defined such that

$$D_{ij} := \begin{cases} d_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

We can now define the Laplacian matrix L of the graph G as $L = D - A$. The Laplacian relates to many useful properties of a graph. Nodes can have embedded features as well, convenient to represent data or node characteristics.

3.1.3 . Modelling cortical surface as a regular mesh

A mesh is a particular type of undirected graph which aims at parcelling an object's surface. It is used to represent continuous surfaces in a discrete manner, the same way an image is represented using pixels on a screen. It is more flexible as it can represent more complicated topologies than regular 2D grids. The cortical surface can be seen as a folded sheet, with most of its surface buried between folds. A spherical template mesh is a natural choice to align subjects brain because it allows to conserve the topological structure of the surface, notably its local connectivity. Two voxels representing neighbouring gyri can be closer in image space than each of them to a voxel representing a part of the in-between sulcus. Yet this property is not desirable considering the biological properties of the cortical surface, because a good cortical surface representation should present tissue continuity properties. In fact, each hemisphere is homotope to a sphere, which can be illustrated using an inflating procedure (as shown in Fig. 3.1).

To properly measure properties of the cortical mantel such as thickness or curvature, expert tools have been specifically developed [86, 120, 32, 119]. They rely on various processing steps to process T1-weighted sMRI and transform it in order to properly measures such proxies of biology. For instance, FreeSurfer [86, 120] includes a dedicated procedure to isolate cortical grey matter from the rest of the brain. It uses meshes to approximate the cortex structure. Preprocessing steps include removal of non-brain tissues, registration in the Talairach space, segmentation of grey / white matter, voxel intensities normalisation following MRI gradients using white matter as reference, tessellation using meshes to represent the borders between white / grey matter and grey matter / CSF. These subject-level tessellations are then inflated to a sphere (one for each hemisphere) and registered to a spherical template defined using shared cortical folding patterns. Throughout these procedures are computed various measures, such as gyrification index, curvature, area or thickness, that are

represented on this spherical mesh.

In parallel, neuroanatomists have been studying the structure of the brain in various ways and proposed many atlases, that are consistent across individual variations, allowing to study the brain within a population and finding biomarkers that are not due to subject-specific variations of anatomy. There are many ways of building an atlas, using ex-vivo or in-vivo brains, with histology techniques or MRI. Some of them are built using cytoarchitecture, others use brain functions, some are defined using white fiber bundles extracted using tractography technique from diffusion MRI, and others defined using sulci morphology. Once an atlas is defined, it can be projected onto this spherical template used by FreeSurfer, to represent the cortical measures averaged over the ROIs defined by the atlas. Freesurfer includes Desikan [94] and Destrieux [96] parcellations, two atlases defined using cortical folding patterns. There is no consensus as to which atlas is the best (for instance Destrieux defines a finer parcellation than Desikan), because it always depends on the application. It is a way to enforce some kind of structure in the data using a prior knowledge.

On the other hand, using atlases might not be the best approach to finding biomarkers of brain pathology. Main arguments regard the lack of precision of these atlases. Indeed, depending on the parcellation, we usually consider coarse average of measures over large areas. While it can smooth some reconstruction related noise in the data, it can also hide small effects that can be relevant to explain pathologies [124]. Others say that it harms reproducibility, since there are so many atlas available that research using one atlas will rarely find the same results as another using a different one [124]. And finally, when using cortical atlases to structure data, the spatial information is not used, since it is usually treated as tabular data. A solution to the aforementioned issues is to handle cortical surfaces at the mesh level. The template spherical mesh used by FreeSurfer is a regular icosahedron of order 7, with approximately 160K vertices per hemisphere. The order of the icosahedron can be reduced easily, modifying resolution without changing its structure: all vertices have 6 neighbours, except for 12 of them which only have 5. This very regular graph structure allows us to define specifically adapted operation such as convolutions, that are able to leverage the mesh structure to increase representational power and reduce computational burden in neural networks for instance.

3.2 . Opportunities for modelling cortical structure using Neural Networks

In this section, we want to support the fact that NNs based models are suited for modelling cortical surface structure. The well-known convolution operator implemented for images has been extended to be able to learn from less restrictive data types, generalising its definition to other geometries such as graphs [40]. In the following, we describe existing graph convolution operators, which are able to leverage cortical surface structure, and their numerous applications in neuroimaging for solving different problems.

3.2.1 . Structure-modelling operators

Graph convolutions and geometric deep learning. Due to the complex nature of a graph, it is not as straightforward as for images to define a convolution operator. A NN implemented using

convolution operator should be equivariant via translation thanks to convolution properties. For 2D or 3D images, translations happen in 2 or 3 directions. In a graph, translations can not be defined the same way, because each node can have a different number of neighbours.
There exist many graph convolutions [40, 90, 198, 351] operators, developed to work with any kind of graph. Some of them [90, 198] rely on graph spectral theory and leverage the graphs Laplacian to perform convolution in Fourier space.
Others [19, 261, 351] use a spatial approach. In particular, [351] uses an attention mechanism (see Section 2.2.2), introduced firstly [350] to represent words within a sentence, to represent nodes with respect to their neighbours.
In the case of regular spherical meshes, when studying registered spherical cortical surface for instance, these approaches are undully computationally expensive because they do not leverage the specificities of the regular icosahedron. Various convolution strategies exist to work with such spherical or icosahedral topologies.

Spherical mesh operators. We will not draw up an exhaustive list of spherical convolution operators but rather to give a brief overview of their properties. More systematic reviews [390] exist in the literature.
To work with spherical mesh specificities, Defferard et al.[91] proposes convolution operators in the spectral domain with interesting properties like rotation equivariance and independence from spherical sampling (i.e. discretising mesh). Similarly, other methods implement convolution with desirable properties for spherical data [70, 110], using spherical harmonics to define rotation equivariant convolutions. But the computational cost is heavy and we argue these properties are not mandatory for our problem, because our data is always registered with the same sampling of the sphere.

The simplest strategy consists in projecting textures on a 2D rectangular planar surface using azimuth and elevation, called EquiRectangular Projection (ERP), and apply a classical 2D convolution with a particular padding for border continuity in the azimuth axis. This introduces deformations that have been shown to harm performances [27, 314]. Others strategies were developed specifically for this type of data by accounting for the regularity of such spherical mesh, avoiding most of these deformations:

- One uses patches [314] of different shape in the tangent plane, including Rectangular Patches (RePa). Using interpolations and projections, a classical convolution can then be used over this patch. Once again, the number of operations makes it not the most efficient and it inevitably introduces deformations.
- Another convolution approach [183] uses differential operators computed over the 1-ring neighbourhood of each vertex to define a convolution on the icosahedral mesh. This definition is interesting for our setting and was applied in previous studies using the cortical surface data [277].
- Finally [394] proposes to order the neighbours according to their angle in the sphere tangent plan on each vertex, and assign them the corresponding index, starting in the middle. On the

regular icosahedron, there are exactly 12 vertices with 5 neighbours, the other all have 6 neighbours, at any order. When a vertex has 5 neighbours, we simply assign the center node indices 0 and 1. This provides an almost always consistent neighbourhood definition and therefore a quite efficient convolution operator, called Direct Neighbour (DiNe).

In our experiments, we will consider the DiNe convolution operator, because it is very efficient while maximally taking advantage of the regular icosahedral mesh. Its definition is close to the classical 2D or 3D image convolution, which makes it easy to improve with further developments [392]. It was found to perform well on a variety of tasks as compared to other graph convolutions and spherical convolutions [113]. You can find an illustrated example of the DiNe convolution in Figure 3.1.

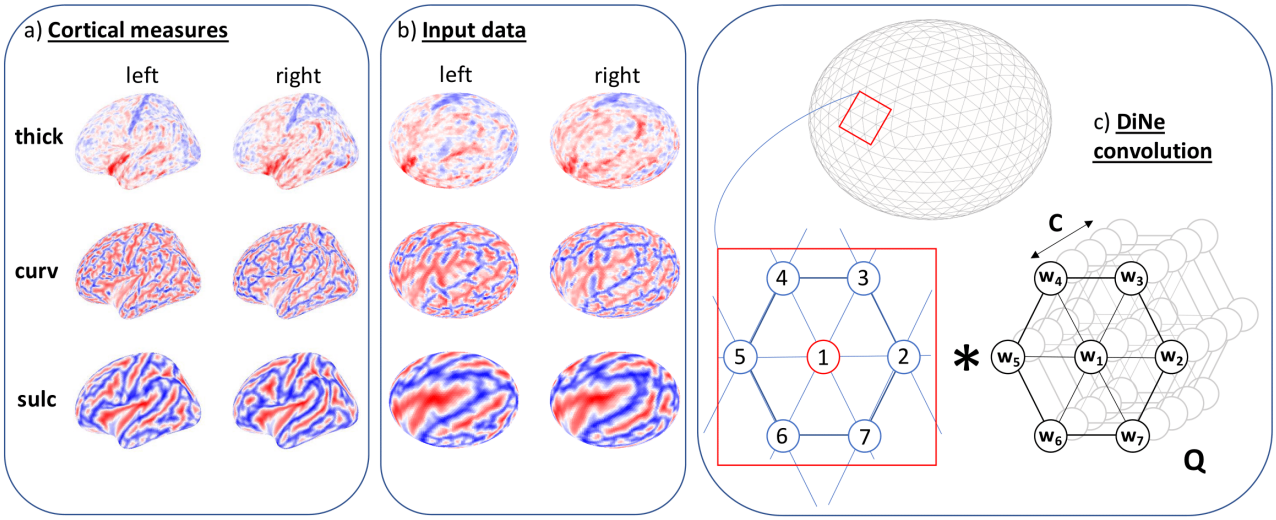


Figure 3.1: Illustration of the Direct Neighbour (DiNe) convolution. a) Measurements of cortical thickness (thick), curvature (curv), and sulcal morphology (sulc) in one subject's left and right hemispheres. Each of these metrics represent a channel ($d = 3$ channels) for each vertex the two surfaces (left and right hemispheres). b) The same hemispheric cortical topologies are then inflated onto a regular icosahedral sphere of order 5. c) Finally, the cortical measures are fed into a DiNe convolution that transforms a d-channel input cortical feature map ($d = 3$ in the first layer) into a C-channel output feature map. Note that the convolution kernel $\boldsymbol{Q}$ contains weights $\boldsymbol{w_i}$.

3.2.2 . Applications modelling cortical surface structure using NNs

Up to today, NNs modelling structure using cortical surface adapted operators as introduced above have been applied to solve a variety of challenges [390]. Interestingly, we can note that most of these applications focus on early processing steps such as surface reconstruction, registration or parcellation, in order to improve classical MRI processing pipeline such as FreeSurfer, by solving these problems faster. Other applications concentrate on phenotype prediction and our work will be more situated in this line of work.

Improving over current surface construction pipelines. FreeSurfer [86, 120] is a widely used tool for sMRI processing, and is composed of various processing steps for cortical surface extraction and registration to a spherical template (fsaverage). It is however relatively slow, with processing times of about 3 hours to extract cortical measures and register them on the fsaverage template. DL based methods have been proposed to accelerate this preprocessing pipeline [162] and these methods fall in this line of literature.

FreeSurfer makes errors when reconstructing the cortical surface, some of which are reflected by the Euler number, which gives a quality control metric for the reconstruction step [301]. This step consist in tessellating the white / grey matter and pial / CSF boundaries. Some methods [79, 367] use 3D volumes directly to reconstruct these surfaces, were others [236, 171, 218, 235] rely on CNNs and mesh deformation methods. However, the latter methods all use FreeSurfer reconstructed surfaces as groundtruth, so they are limited by FreeSurfer performances.

Cortical registration consists in smoothly mapping reconstructed surfaces to a template. The fsaverage spherical template using a regular icosahedral sampling is commonly used. Early work [62] using DL for cortical surface registration used 2D ERP projection, introducing unevitable deformations. More advanced methods use spherical adapted operators such as DiNe to solve this problem [391] and latest work [393] propose to solve this jointly with cortical surface parcellation.

Cortical parcellation consists in attributing atlas labels to reconstructed cortical surfaces. It is very related to cortical registration as labels are usually attributed to vertices on the spherical template before being propagated back to the reconstructed surfaces. An early NN based method proposes to compute 2D convolution over tangent patches to the spherical surface to assign Desikan labels [373] to each of these non-registered vertices. Leveraging the regular icosahedral sampling of the fsaverage template, the Spherical U-Net [394] introduces the DiNe convolution operator to build cortical parcellation of the child brains. Another approach using convolution over a 1-ring neighbourhood for cortical surface parcellation was proposed [277]. It exploits the spherical structure of the surface by employing a differential spherical convolution operator [183]. Other approaches [33, 80, 109, 143, 374, 225] use general graph CNN to avoid relying on a previous registration step.

Phenotype prediction. On the other hand, cortical surface data has been exploited for phenotype prediction or longitudinal modelling. The latter consist in modelling the evolution of structural or functional parameters (e.g. curavture, thickness and functional connectivity) of the cortical surface. DL based methods modelling structure through convolutional operator were shown to better predict changes in cortical morphology / function than their conventional machine learning counterparts [231, 394, 267, 291]. Other prediction tasks include age, sex and other cognitive capacity such as Full Scale Intelligence Quotient (FSIQ). Finding aging biomarkers is of interest in order to understand the development process of the brain and related brain conditions such as Schizophrenia and Alzheimer's diseases [71]. Seong et al. [314] predicts sex from cortical surface introducing the RePa convolution operator, where other works [230, 61, 85] additionally try to find cortical biomarkers by predicting age or cognitive phenotypes using graph CNNs. There is

no consensus as which strategy is the best between using general graph convolution or specific spherical operators such as DiNe [113] for predicting different phenotypes, even if spatial convolution filters appear to perform better than their spectral counterparts. Regarding prediction of condition status, most approaches using spherical or graph convolutions for cortical surface modelling [262, 116, 27] have focused on Alzheimer's disease prediction. Another approach [142] proposes a general graph adaptive pooling operator and predicts diseases status from cortical surface and regresses age.

All these applications illustrate the pertinence and available tools to leverage cortical structure from cortical surface data. It can be presented either directly on brain reconstructed surfaces or registered onto a spherical template. Having a spherical template allows ML models to find patterns and biomarkers of the cortical surface consistent across individuals. Newly developed graph or spherical specific convolution operators allow to properly model the cortical structure represented on this template and have demonstrated good prediction performances. However, these methods have only been applied in supervised prediction settings, relying on manually labels data examples, which is a very limiting setting in neuroimaging and more generally in medical imaging. Self-supervised learning methods are being developed (see Section 2.2.4) to properly learn from unannotated data. Their extension to cortical surface data requires a few adjustment, including the definition of cortical surface data augmentations.

3.3 . Self-supervised learning on cortical surface

Most applications using cortical surface are prediction or segmentation supervised tasks. The development of efficient icosahedral operator for NNs and advances in unsupervised – and more particularly self-supervised – learning uncovers new opportunities for exploiting cortical surface data.

Data-driven studies of brain pathologies are often hampered by the scarcity of available data, leading to potential failures in the discovery of statistically significant biomarkers. Key factors include recruitment of rare disease patients and acquisition costs. Many research efforts have attempted to address this challenge [107], but it has been shown that finding reproducible biomarkers requires a large amount of data [240], an accentuated effect when using DL techniques [103]. Transfer learning has become a promising solution with the advent of large cohorts such as OpenBHB [105]. Transfer learning consists of training a neural network with pretext tasks on a large dataset. The trained NN is then fine-tuned on a smaller, application-specific dataset. However, transfer learning for medical imaging is still in its early stages. Interestingly, there is a consensus that using a natural image dataset may not lead to the best transfer strategy [286, 8].

In recent years, several training schemes have been proposed for learning "universal" data representations [36]. The goal is to summarise as much of the semantic information as possible. Among the most promising approaches are self-supervised schemes that can provide good NN initialisation for transfer learning [49, 147, 384]. Specifically, contrastive learning uses data augmentation to structure the learned latent space [60, 158, 194]. Almost all SSL techniques rely heavily on the data augmentation [49, 60, 147, 383]. Currently, data augmentation for medical imaging is only available

for image data defined on a regular rectangular grid.

In this section, we refer to domain-specific architectures by defining Spherical Convolutional Neural Network, that use such specific convolution operators introduced in Section 3.2.1, and in particular the DiNe operator. To train SCNNs using SSL, we introduce three baseline and two original augmentations specifically designed for the brain cortical surfaces (Fig. 3.3). The three baseline augmentations are directly inspired by natural image transformations: the SurfCutOut involves cutting out surface patches, the SurfNoise adds Gaussian noise at each vertex, and the SurfBlur applies Gaussian blur. The proposed MixUp augmentations build on the original idea of randomly selecting some cortical measures and replacing them with realistic corrupted samples: the HemiMixUp exploits the symmetry of the brain and permutes measures between hemispheres of the same individual, and the GroupMixUp bootstraps vertex-based measures from a group of similar individuals. We illustrate how these augmentations can fit into the well-known and illustrating SimCLR [60] self-supervised scheme. We provide a comprehensive analysis to investigate the learned cortical surface representation using self-supervised learning with SCNNs and the proposed augmentations. The representations are evaluated using different downstream tasks, which consist of predicting age, sex and a cognitive phenotype. The experimental setting is illustrated in Fig. 3.2.

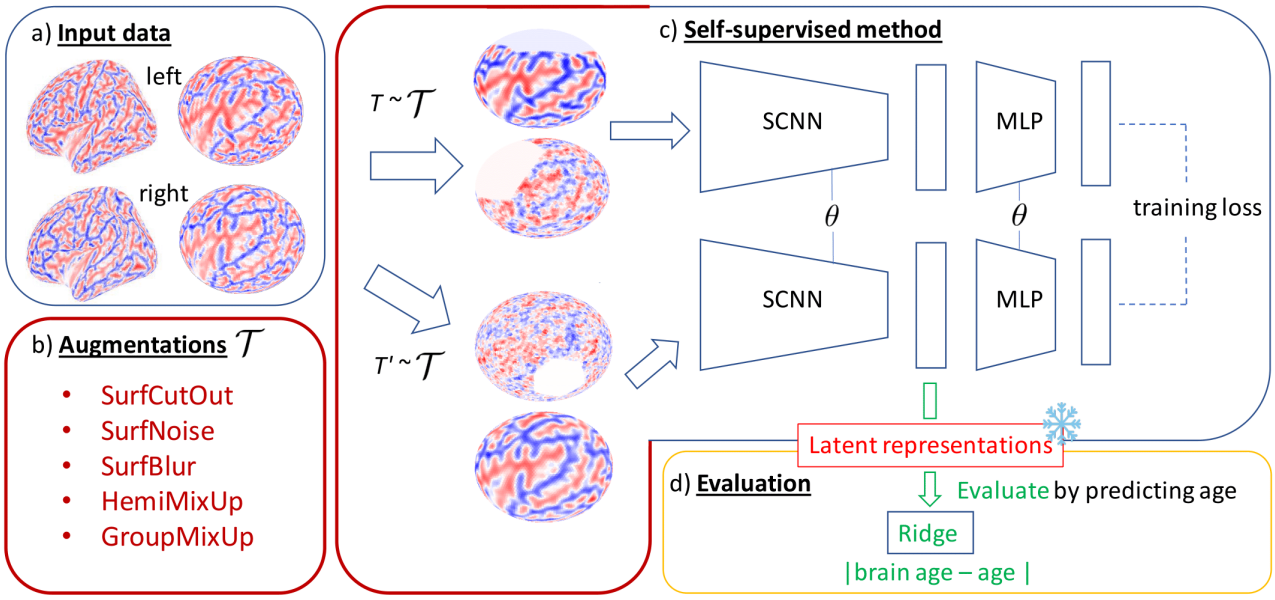


Figure 3.2: Overview of the proposed evaluation framework for spherical augmentations: a) input cortical measures (here curvature) inflated to a sphere, b) a set of adapted and domain-specific augmentations $\mathcal{T}$ that allow the generation of augmented cortical measures, c) a self-supervised model with parameters θ consisting of a Spherical Convolutional Neural Network (SCNN) and a MultiLayer Perceptron (MLP) projector, and d) evaluation of the model's frozen representations using a linear predictor (here the mean absolute deviation between linearly predicted brain age and true age).

3.3.1 . Contrastive Learning

Constrastive learning is a recent SSL scheme, able to learn representations with high semantic content, very generalisable and transferable. It is classified as a deep metric learning SSL algorithm [24]. It relies on the principle stating that two objects with similar (or dissimilar) semantic should be close (or far appart) in representational space, and uses the deep metric learning loss infoNCE to enforce it [325, 323, 273].
There are different way of defining semantic. In a supervised setting, the sole semantic we try to capture is label related. In this sense, two data points with same label are semantically identical. We can see how this definition is quite limiting. In an unsupervised setting, the definition of semantic is more complex. It depends on the considered object. A trivial assumption is that an object has the same semantic as itself. With this in mind, we can try to apply semantically preserving transformations (i.e. augmentations) that can modify a data example without changing its semantic content. In natural images, we know some invariance properties, that can lead us to easily define augmentations to produce modified version of an image without changing its semantic too much.

In particular, contrastive learning as gained in popularity with SimCLR [60], offering a SSL paradigm that relies on the InfoNCE loss and a range of classical data augmentations (e.g. crop, cutout, color distortion or noise). Using a set of stochastic transformations $\mathcal{T}$, the network tries to bring closer in representation space two augmented views $t(x)$ and $t'(x)$, $t, t' \sim \mathcal{T}$ of the same example x, while repelling their representation from that of other augmented examples.
Let f be the encoder network, and g a non-linear MLP, hereafter referred to as projector. We note $z = g(f(x))$ the output embeddings. We randomly sample a minibatch with n examples. Each example is augmented twice using transformations $t, t' \sim \mathcal{T}$, providing n positive pairs among $2n$ examples. Negative pairs correspond to every other pairs within the minibatch. This is the InfoNCE objective optimised by the network:

$$l(i,j) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{\{k \neq i\}} \exp(\text{sim}(z_i, z_k)/\tau)} \tag{3.1}$$

where $\mathbb{1}_{\{k \neq i\}} \in \{0, 1\}$ is the indicator function evaluating to 1 iff $k \neq i$, τ a temperature parameter, and sim a similarity function, usually defined as the normalised scalar product, i.e. $\text{sim}(z_i, z_j) = z_i^\intercal z_j / (\|z_i\| \|z_j\|)$. In particular, we compute this loss for every positive pair (i, j) and (j, i) of the minibatch. Intuitively, we can see that the numerator of this loss will try to bring closer together embeddings coming from augmented view from the same examples, while the denominator will repel them from others, in representation space. In practice, such contrastive approaches are very reliant on the size of the dataset and hardware, because its performances scale with batch size.

In order to use such SSL algorithms, we need to define sets of transformations $\mathcal{T}$ that preserve cortical surface semantic, while offering a challenging contrastive prediction task.

3.3.2 . Augmentations on the Cortical Surface

We propose two types of augmentations. Baseline augmentations refer to augmentations that were derived from similar augmentation defined and commonly used on rectangular grid data such as natural images. We propose to use an equivalent formulation on cortical surface data. Additionally, we propose MixUp augmentations, specifically developped for cortical surface data, where some

randomly selected vertices and replaced by noisy values. We introduce two strategies to implement MixUp augmentations.

Baseline augmentations. All augmentations defined for natural and medical images are not directly applicable to the cortical surface. In a self-supervised scheme, an effective augmentation should reflect invariances that we want to enforce in our representation. It is not a requirement that these augmentations produce realistic samples. Their goal is to provide synthetic constrastive hard prediction tasks [60, 345]. They must also be computationally efficient. For example, geometric transformations such as cropping, flipping, or jittering cannot be applied to cortical measures. We could use small rotations as proposed in [392], but such an augmentation is not computationally efficient due to multiple interpolations and less effective in nonlinear registration cases. We adapt three baseline domain-specific transformations consisting of cutting out surface patches (**SurfCutOut**), adding Gaussian noise at each vertex (**SurfNoise**), and Gaussian blurring (**SurfBlur**). Specifically, the SurfCutOut sets an adaptive neighbourhood around a random vertex to zero. The neighbourhood is defined by R concentric rings of vertices (for the definition of a ring, see the previously introduced DiNe operator in Section 3.2.1). On structural MRI images, a cutout strategy has proven its efficiency in a similar contrastive learning setting [104]. Then, the SurfNoise adds a Gaussian white noise with standard deviation σ_1 (to weight the signal-to-noise ratio), and the SurfBlur smooths the data by applying a Gaussian kernel with standard deviation σ_2 (which controls the spatial extent expressed in rings).

Proposed MixUp brain cortical augmentations. Here, we assume that the structural measures across the cortical surface have a vertex-to-vertex correspondence for both hemispheres. We propose to randomly select vertices and their associated cortical measures and to replace them with noisy realistic ones. A similar approach has been proposed by Yoon et al. [379] for tabular data, and a comparable augmentation has been used for natural images in supervised contexts by mixing up labels [386]. All augmentations are applied on a cortical measure and hemisphere basis. A corrupted version $\tilde{\mathbf{x}}$ of a cortical measure $\mathbf{x} \in \mathbb{R}^P$, where P is the number of vertices, is generated as follows:

$$\tilde{\mathbf{x}} = \mathbf{m} \odot \bar{\mathbf{x}} + (1 - \mathbf{m}) \odot \mathbf{x} \tag{3.2}$$

where $\odot$ is the point-wise multiplication operator, $\bar{\mathbf{x}}$ is a noisy sample to be defined, and $\mathbf{m} \in \{0, 1\}^P$ is a binary random mask. In our case $\mathbf{x} \in \{x_1, \ldots, x_N\}$, where N is the number of subjects. The proposed HemiMixUp and GroupMixUp augmentations offer different ways to construct $\bar{\mathbf{x}}$ (Fig. 3.3). The mask $\mathbf{m}$ is generated by drawing binary random numbers from a Bernoulli distribution $\mathcal{B}(p)$, where p is a hyperparameter controlling the proportion of $\mathbf{x}$ to be modified. In both cases, the augmentation is done at the subject level.

HemiMixUp. This augmentation randomly permutes a subject's measurements at specific vertices across hemispheres, assuming a vertex-to-vertex correspondence between hemispheres. Considering the left hemisphere, we get:

$$\tilde{\mathbf{x}}_i^{\text{left}} = \mathbf{m} \odot \mathbf{x}_i^{\text{right}} + (1 - \mathbf{m}) \odot \mathbf{x}_i^{\text{left}} \tag{3.3}$$

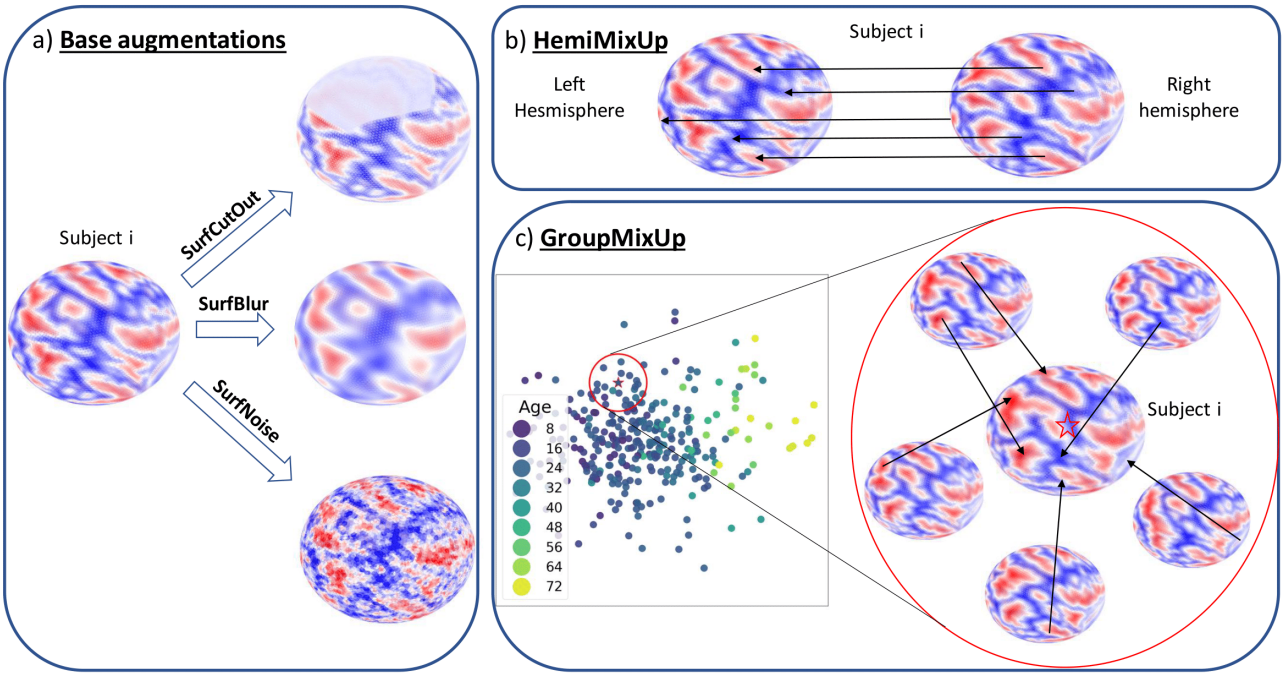


Figure 3.3: Illustration of the considered cortical augmentations: a) the three baselines, b) the proposed HemiMixUp, and c) the proposed GroupMixUp brain cortical augmentations. In the GroupMixUp, groups are defined from the reduced input data using PCA embeddings colored by age. The arrows in b) and c) represent how the cortical measures are modified and are explained in more detail in 3.4.

where $\mathbf{x}_i^{\text{left}}$ and $\mathbf{x}_i^{\text{right}}$ are the left and right hemisphere measures of the subject i, respectively.

GroupMixUp. The GroupMixUp augmentation randomly bootstraps measures at specific vertices across a group of K subjects $G_i = \{g_1, \dots, g_K\}$ sharing similar cortical patterns with respect to the i-th subject. We aim to generate realistic noisy measures without missing hemispheric asymmetries by exploiting the group variability. We define $\mathbf{X}_i = (\mathbf{x}_{g_1}, \dots, \mathbf{x}_{g_K})^\intercal \in \mathbb{R}^{K \times P}$. Considering the left hemisphere, we get:

$$\tilde{\mathbf{x}}_i^{\text{left}} = \mathbf{m} \odot \text{diag}\left[MX_i^{\text{left}}\right] + (1 - \mathbf{m}) \odot \mathbf{x}_i^{\text{left}} \tag{3.4}$$

where diag is the diagonal operator, and $M \in \{0,1\}^{P \times K}$ is the random selection matrix. Each row of M selects a particular subject and is generated by drawing a random location from a uniform distribution $\mathcal{U}(1, K)$. The G_i grouping relies on a PCA trained on the residualised input data. The residualisation is performed using ComBat to remove unwanted site-related noise [121]. We use a K-Nearest neighbours (with Euclidean distance) in the PCA space to define the group G_i. This step is performed only once before training, with little computational overhead. It is important to note that this strategy builds groups from a semantically meaningful space which maximises the explained variance of the data. The first PCA axis is strongly related to age, as shown in Figure 3.3-c). Groups

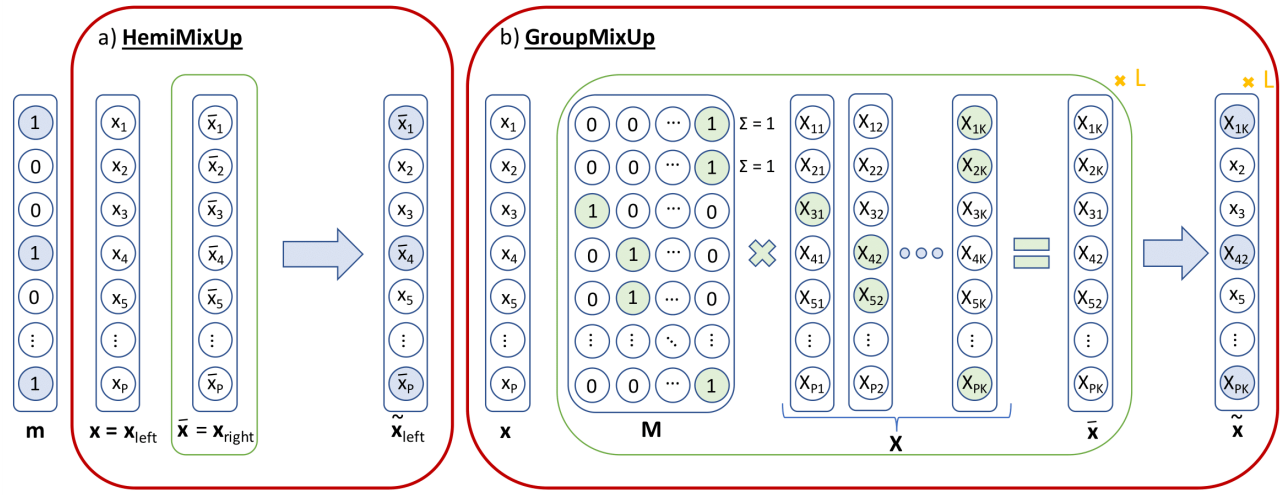


Figure 3.4: Schematic illustration of the a) HemiMixUp and b) GroupMixUp augmentations (single measure $d = 1$ example). From Eq. 1, we generate a corrupted version $\tilde{\mathbf{x}}$ of a cortical measure $\mathbf{x} \in \mathbb{R}^P$, with P the number of vertices, by using a binary random mask $\mathbf{m} \in \{0,1\}^P$, and a realistic noisy sample $\bar{\mathbf{x}}$. The generation of $\bar{\mathbf{x}}$ is illustrated in the green frames. a) The HemiMixUp augmentation defines $\bar{\mathbf{x}}$ as the controlateral cortical measures assuming a vertex-to-vertex correspondence between hemispheres. b) The GroupMixUp augmentation defines $\bar{\mathbf{x}}$ from a set of K similar cortical measures grouped in $\mathbf{X} \in \mathbb{R}^{K \times P}$. The random selection matrix $M \in \{0,1\}^{P \times K}$ allows the generation of a realistic noisy sample $\bar{\mathbf{x}}$. With the GroupMixUp augmentation, we can generate L realistic noisy samples from the same subject by drawing different selection matrices M.

are formed independently for each individual's hemisphere.

By inverting the *left* and *right* notations in equations 3.3 and 3.4, the formulations hold for the right hemisphere. These two MixUp augmentations are further illustrated in Figure 3.4.

3.4 . Experiments and validation

This section introduces the implemented experimental setting to validate our SSL approach on cortical surface data, leveraging our proposed data augmentations and specific SCNNs architectures. More precisely, we start by describing the datasets used, the selected SSL scheme, how we select and evaluate the models and how we compare the proposed approach to classical supervised approaches using the same SCNN architectures. We then discuss the obtained results and their implications.

3.4.1 . Data and settings

Datasets. The T1-weighted structural MRI data are processed with FreeSurfer, which calculates thickness, curvature, and sulcal morphology for each cortical vertex [120]. Interhemispheric registration (XHemi) is performed to obtain vertex-to-vertex mapping between hemispheres [146]. Inflated hemispheric cortical topologies are finally expressed on a regular order-5 icosahedral sphere (around 10K vertices). We use two datasets to demonstrate the proposed augmentations. First, we use the

OpenBHB, including more than 5000 individuals (age distribution 25.3 ± 15.0) with images coming from multiple acquisition sites [105]. The cortical surface extracted measures and registered on the fsaverage are available for all the participants, using the Euler number as Quality Control (QC), keeping those with Euler < -217 [301]. We use the so-called OpenBHB internal train and test sets (with the same imaging acquisition sites). We further split the OpenBHB internal test set into validation and test sets (hereafter referred to as internal test), preserving the population statistics (age, sex, and acquisition site). Finally, we keep unchanged the OpenBHB external set (hereafter referred to as external test), which consists of subjects with a similar age distribution but disjoint acquisition sites. This set is used to evaluate generalisation and robustness to unseen sites. Second, the Healthy Brain Network (HBN) cohort, which includes more than 1700 children (age distribution 10.95 ± 3.43) with behavioral specificities or learning problems [6]. After applying the same quality control as for OpenBHB images, we keep 1407 subjects. We split the data into training (80 %) and test (20 %) sets, preserving population statistics (age, sex and acquisition site). Some subjects (1073) have the cognitive score WISC-V FSIQ available.

The self-supervised model. SimCLR [60] contrastive learning strategy attempts to bring two representations of the same transformed sample as close as possible in the model latent space, while repelling other representations (more details in Section 3.3.1). We implement this model following the recent literature on self-supervised learning, which consists of an encoder and a projector. For the encoder, we choose a single SCNN architecture to facilitate the comparison between the methods (see Appendix Tab. A.1). It has four convolution blocks. Each convolution block consists of a DiNe convolution layer followed by a rectified linear unit and an average pooling operator [394]. There are two branches in the first convolution block, one for each hemisphere. The resulting features are concatenated on the channel axis and piped to the network flow. For the projector, we implement the architecture recommended in [60], a MultiLayer Perceptron (MLP). The model is trained on the OpenBHB training set. Three augmentation combinations are considered during training: all the baseline brain cortical augmentations (Base), Base + HemiMixUp, and Base + GroupMixUp. The entire procedure is repeated three times for each combination (nine trainings in total) to obtain a standard deviation for each prediction task described below.

Model selection and evaluation. In self-supervised learning, the training loss, even when evaluated on a validation set, indicates convergence but does not reflect the quality of the learned representations [359, 307]. As suggested in the literature to overcome this problem, we add a machine learning linear predictor on top of the encoder latent representations during training (ridge for regression and logistic for classification) (Fig. 3.2). We then estimate and monitor the associated prediction score from the validation set at each epoch (Mean Absolute Error (MAE) and coefficient of determination R^2 for regression and Balanced Accuracy (BAcc) for classification). This score is only used to monitor the training, leaving the SimCLR training process completely unsupervised. Most network, training and augmentation hyperparameters (see Appendix Tab. A.1 and Tab A.2) are the same and were fixed by testing several values for age prediction. Exceptions are the number of epochs and the l_2 regularisation strength of the ridge and logistic predictors, which were set by using the evaluation criteria corresponding to the final task. Finally, we evaluate the trained models with the same

strategy for age and sex predictions on all cohorts and for FSIQ prediction on HBN. Age, sex and FSIQ are known to be proxy measures to investigate mental health [84]. They represent features that a pre-trained self-supervised model should be able to learn and generalise to unseen data. Due to the discrepancy in age distribution and the lack of clinical variables of interest, the linear predictors are fitted to the OpenBHB and HBN training representations. Finally, the OpenBHB internal and external test sets and the HBN test set are used to evaluate the prediction scores.

Model comparison. We compare the proposed SimCLR-SCNN model with supervised SCNNs. The supervised models consist of the same encoder followed by a linear predictor. The training loss for these supervised models depends on the task at hand (l_1 for regression and cross-entropy for classification). These supervised models will be referred to as age-supervised if they were trained to predict age, and sex-supervised if they were trained to predict sex. Each supervised model is trained 3 times as well, to derive standard deviations, on the same train set as self-supervised models. They are evaluated the same way as self-supervised models: task-dependent machine learning linear predictors (ridge for regression and logistic for classification) are fitted to the learned representations from the SCNN encoder and evaluated on the test representations.

3.4.2 . Results

Self-supervised SCNNs generalise better than supervised SCNNs. On OpenBHB, compared to a much more specialised supervised SCNN setup, a SCNN trained with the SimCLR self-supervised learning framework and the proposed augmentations shows a rather comparable performance for each of the investigated tasks (Table 3.1-a). For example, on the internal test, the SimCLR-SCNN age-MAE scores are 4.87, 4.72, and 4.55 for the Base, Base + HemiMixUp and Base + GroupMixUp augmentations, respectively. These scores are slightly worse than the predictions of the age-supervised SCNN (4.0 age-MAE), which is expected since the latter was trained in a supervised manner to learn good representations for predicting age. However they remain comparable and largely outperform the sex-supervised SCNN (6.2 age-MAE). The same trend can be observed for the R^2 for the age prediction task and the BAcc for the sex prediction task for all test sets. Remarkably, for some tasks, the SimCLR-SCNN with the proposed augmentations even outperforms supervised SCNN models in terms of generalisation performance. This can be seen by comparing the results on OpenBHB internal and external test sets. For example SimCLR-SCNN loses 10% ($0.81 \rightarrow 0.71$), 10% ($0.81 \rightarrow 0.71$), and 8% ($0.82 \rightarrow 0.74$) BAcc for the Base, Base + HemiMixUp and Base + GroupMixUp augmentations, between internal and external test sets, while age- and sex-supervised SCNNs lose 12% ($0.67 \rightarrow 0.55$) and 18% ($0.86 \rightarrow 0.68$) BAcc respectively. SimCLR-SCNNs even outperform sex-supervised SCNNs for the sex prediction task on the external test set. As expected, the prediction on the external test decreases for both strategies. When the learned knowledge is transferred to HBN, we show better (or at least equivalent for sex) prediction performance for the SimCLR-SCNN with the proposed augmentations compared to the supervised SCNNs. Note that the age distribution in HBN is much narrower with a younger population than in OpenBHB. Therefore, the MAE between Tables 3.1-a and 3.1-b cannot be directly compared. Interestingly, the R^2 values are still comparable and show a stable goodness of fit for the SimCLR-SCNN model. Note that using a supervised MLP with more than 120M parameters to predict age from the same input data gives only slightly better results than using the SimCLR-SCNN

a)

SimCLR-SCNN	**OpenBHB internal test**			**OpenBHB external test**		
Augmentations	Age		Sex	Age		Sex
	MAE (↓)	R^2 (↑)	BAcc (↑)	MAE(↓)	R^2 (↑)	BAcc (↑)
Base	$4.87_{\pm 0.14}$	$0.81_{\pm 0.01}$	$0.81_{\pm 0.01}$	$5.89_{\pm 0.17}$	$0.50_{\pm 0.03}$	$0.71_{\pm 0.01}$
Base + HemiMixUp	$4.72_{\pm 0.16}$	$0.83_{\pm 0.01}$	$0.81_{\pm 0.01}$	$5.62_{\pm 0.13}$	$0.55_{\pm 0.03}$	$0.71_{\pm 0.02}$
Base + GroupMixUp	$\mathbf{4.55_{\pm 0.07}}$	$\mathbf{0.84_{\pm 0.01}}$	$\mathbf{0.82_{\pm 0.01}}$	$\mathbf{5.47_{\pm 0.15}}$	$\mathbf{0.58_{\pm 0.02}}$	$\mathbf{0.74_{\pm 0.01}}$
Supervised-SCNN						
Age-supervised	$4.00_{\pm 0.12}$	$0.84_{\pm 0.01}$	$0.67_{\pm 0.01}$	$5.06_{\pm 0.19}$	$0.61_{\pm 0.02}$	$0.55_{\pm 0.01}$
Sex-supervised	$6.20_{\pm 0.20}$	$0.69_{\pm 0.02}$	$0.86_{\pm 0.01}$	$6.40_{\pm 0.32}$	$0.42_{\pm 0.06}$	$0.68_{\pm 0.01}$

b)

SimCLR-SCNN	**HBN test**				
Augmentations	Age		Sex	FSIQ	
	MAE (↓)	R^2 (↑)	BAcc (↑)	MAE(↓)	R^2 (↑)
Base	$1.69_{\pm 0.07}$	$0.65_{\pm 0.02}$	$0.80_{\pm 0.01}$	$12.97_{\pm 0.25}$	$0.10_{\pm 0.02}$
Base + HemiMixUp	$1.68_{\pm 0.02}$	$\mathbf{0.66_{\pm 0.01}}$	$0.80_{\pm 0.01}$	$\mathbf{12.71_{\pm 0.16}}$	$\mathbf{0.13_{\pm 0.02}}$
Base + GroupMixUp	$\mathbf{1.66_{\pm 0.02}}$	$\mathbf{0.66_{\pm 0.01}}$	$\mathbf{0.81_{\pm 0.01}}$	$\mathbf{12.60_{\pm 0.29}}$	$\mathbf{0.14_{\pm 0.03}}$
Supervised-SCNN					
Age-supervised	$1.74_{\pm 0.05}$	$0.63_{\pm 0.02}$	$0.66_{\pm 0.01}$	$13.05_{\pm 0.06}$	$0.09_{\pm 0.002}$
Sex-supervised	$1.72_{\pm 0.01}$	$0.63_{\pm 0.01}$	$0.82_{\pm 0.0048}$	$12.79_{\pm 0.01}$	$0.09_{\pm 0.01}$

Table 3.1: Evaluation of the learned representations using a machine learning linear predictor on different OpenBHB (a)/HBN (b) sets of data, tasks, and metrics. The proposed MixUp augmentations (HemiMixUp and GroupMixUp) are evaluated against combined baseline (Base) augmentations (SurfCutOut, SurfBlur and SurfNoise) using an unsupervised SimCLR-SCNN framework. The results are compared to supervised SCNNs trained to predict either age or sex (see section **Model Comparison** for details).

($\sim$ 2M parameters) trained with the Base + GroupMixUp augmentations (4.85 *vs* 5.47 age-MAE on the OpenBHB external test) [105]. This suggests that the SimCLR-SCNN model with the proposed augmentations is able to learn good representations without supervision and without being too much biased by the acquisition site.

The MixUp augmentations improve performance. It is clear that the MixUp augmentations improve the learned representations for each prediction task of the OpenBHB internal and external tests (Table 3.1-a). In practice, we found that the GroupMixUp works better than the HemiMixUp augmentation strategy. This can be explained by the attenuation of some properties of the interhemispheric asymmetry forced by the HemiMixUp augmentation. Although the improvement in predicting age and sex on HBN test set is inconclusive, it is clear that HemiMixUp and GroupMixUp help in predicting FSIQ, especially when looking at the R^2 metric (Table 3.1-b).

3.4.3 . Ablation Study: which Base augmentation is the most relevant ?

a)

SimCLR-SCNN	**OpenBHB internal test**			**OpenBHB external test**		
Augmentations	Age		Sex	Age		Sex
	MAE (↓)	R^2 (↑)	BAcc (↑)	MAE(↓)	R^2 (↑)	BAcc (↑)
SurfBlur	$5.79_{\pm 0.07}$	$0.71_{\pm 0.00}$	$0.70_{\pm 0.01}$	$7.09_{\pm 0.54}$	$0.34_{\pm 0.09}$	$0.61_{\pm 0.01}$
SurfCutOut	$5.25_{\pm 0.06}$	$0.78_{\pm 0.004}$	$0.80_{\pm 0.002}$	$6.49_{\pm 0.28}$	$0.43_{\pm 0.03}$	$0.71_{\pm 0.02}$
SurfNoise	$5.30_{\pm 0.03}$	$0.77_{\pm 0.01}$	$0.81_{\pm 0.01}$	$6.32_{\pm 0.13}$	$0.46_{\pm 0.02}$	$0.70_{\pm 0.01}$

b)

SimCLR-SCNN	**HBN test**		
Augmentations	Age		Sex
	MAE (↓)	R^2 (↑)	BAcc (↑)
SurfBlur	$1.85_{\pm 0.01}$	$0.58_{\pm 0.01}$	$0.70_{\pm 0.01}$
SurfCutOut	$1.72_{\pm 0.03}$	$0.63_{\pm 0.02}$	$0.78_{\pm 0.02}$
SurfNoise	$1.76_{\pm 0.02}$	$0.60_{\pm 0.001}$	$0.77_{\pm 0.0046}$

Table 3.2: Additional ablation study results: evaluation of the learned representations using a machine learning linear predictor on different OpenBHB (a)/HBN (b) sets of data, tasks, and metrics. The proposed baseline augmentations (SurfCutOut, SurfBlur and SurfNoise) are evaluated against each other using the unsupervised SimCLR-SCNN framework.

We provide additional results reporting downstream performances of predicting either age or sex, of our SimCLR-SCNN model when it was trained using each Base augmentation individually. These experiments are led to assess which of these augmentations are the most suited for learning cortical surface representation using a SSL schemes, which is also highly related to which of the inductive bias they implement is best to learn representations from cortical surface data. We aggregate the results in Table 3.2. From Table 3.2-a), we can see that alone, SurfBlur seems to be less relevant than SurfCutOut and SurfNoise for good downstream performances overall. For instance, on internal OpenBHB, SimCLR-SCNN age-R^2 are 0.71, 0.78 and 0.80 for SurfBlur, SurfCutOut and SurfNoise respectively, or for sex prediction 0.70, 0.80, 0.81 BAcc respectively. This can be seen for OpenBHB external test and HBN test sets as well.
However, comparing SurfCutOut with SurfNoise is less evident. On OpenBHB internal, training with solely SurfCutOut as augmentation seems to give better or similarly performing representations for predicting age than training with SurfNoise. Indeed, for age-MAE on internal OpenBHB, SimCLR-SCNN achieves 5.25 and 5.30 when training with SurfCutOut or SurfNoise respectively. This is holds when comparing their prediction performances on HBN test set as well (e.g. age-R^2 0.63 and 0.60 respectively). However, when comparing them on OpenBHB external test, SimCLR-SCNN achieves 6.49 and 6.32 age-MAE when trained with SurfCutout or SurNoise respectively. For the sex prediction task, they perform comparably on every test set.

Overall, each one of these three augmentations can be used alone to train our SimCLR-SCNN model and give reasonably good downstream performance, but SurfCutOut and SurfNoise seem more suited than SurfBlur to learn good and generalisable cortical surface representations. This can be due to the fact that blurring is less local than cuting out patches of surface or adding noise sporadically, hence erasing some relevant information. However, using them together will always lead to better and more generalisable representations. For instance, comparing age-MAE on internal OpenBHB 4.87 and 5.25 for SimCLR-SCNN trained with all three Base augmentations (in Table 3.1) or solely with SurfNoise (in Table 3.2) respectively. Comparing age-R^2 on HBN test set shows 0.65 and 0.63 for SimCLR-SCNN trained with all three Base augmentations (in Table 3.1) or solely with SurfCutOut (in Table 3.2). For the sex prediction task, the same observation can be made, yet with less gap: SimCLR-SCNN trained with all three Base augmentations (in Table 3.1) is only significantly better than when it is trained solely with SurfCutOut (in Table 3.2) for sex prediction on HBN test set, where they perform 0.80 and 0.78 BAcc respectively. Overall, the best strategy remains to combine all these Base augmentations together, which is the same as for natural images [60]. Taken individually, SurfCutOut or SurfNoise alone can be used to learn relatively good and generalisable cortical surface representations.

3.5 . Open source developments

3.5.1 . An open source library: surfify

Surfify[1] is an open source module allowing to easily build models and training pipelines able to learn from icosahedral cortical surfaces data output from FreeSurfer [120] for instance. It implements many usefull tools for general icosahedral mesh manipulations, such as augmentations, and building specialised network architectures with a few convolution kernels selected from the literature. Some classical NN architectures are implemented as well. Augmentations introduced in Section 3.3.2 are implemented in the repository. This module was presented during OHBM 2022 at Glasgow during poster sessions. It is build upon various open source widely used python packages, described in Figure 3.5.

Convolutions

The module implements three convolutions strategies. The first one consists of projecting the surface onto a 2D planar rectangular surface such as a World's map, previously introduced as ERP. Then any classical 2D image convolution can be used. A circular padding operator can be used jointly in order to keep proximity between borders, which are in fact neighbours on the spherical mesh. This strategy, called Spherical Mapping (SpMa) in surfify, is simple but it introduces some deformations, which have been shown to deterior performance in various downstream tasks [314, 27].
The two other implemented convolutions were specifically developed for data with this type of structure. Rectangular Patching (RePa) [314] can be used in the tangential plan of each vertex to project the surrounding mesh vertices on it. Then a 2D convolution is applied on this patch. Direct Neighbour (DiNe) [394] consists of taking full advantage of the regular nature of the input icosahedron, and is

[1]Link to the github repository

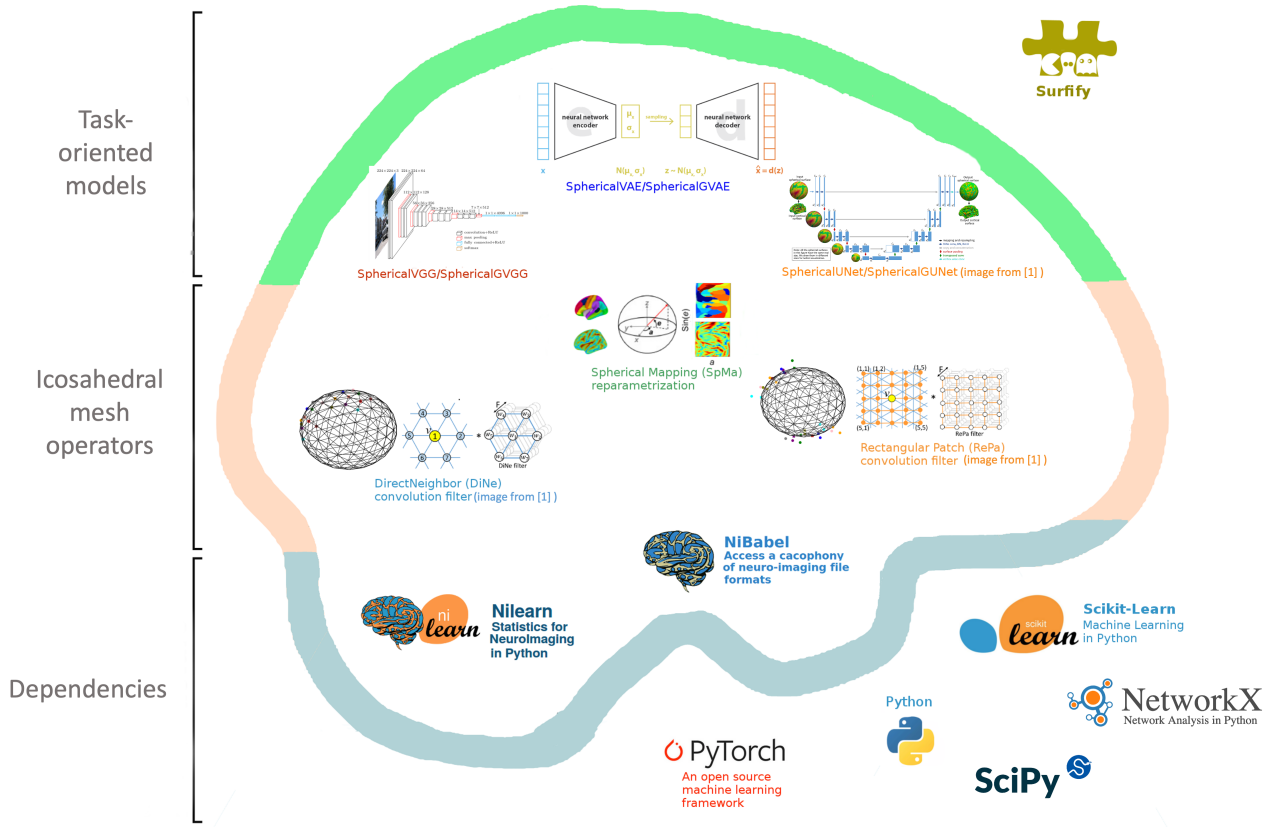


Figure 3.5: Illustration of the open source framework surfify.

described precisely in 3.2.1.
These convolution operators come with pooling operators, allowing the network to reduce the dimensionality of the data progressively. On the icosahedral mesh, it consists mostly in reducing the order of the icosahedron each time, by aggregating the values of neighbouring vertices to their corresponding centre vertex, using pooling function such as mean or max [314, 394].

Augmentations

In addition to the five augmentations introduced in the Section 3.3.2, we implemented an additional SurfRotation augmentation to perform rotations. It uses interpolations, which makes it less efficient than other augmentations, so we did not use it in the study introduced in Section 3.4, as SSL is usually performed on big datasets and requires fast and efficient transformations as augmentations. Execution time on a single metric displayed on icosahedron at different orders are provided in Table 3.3.

Networks

We implemented a few classical NN architectures to demonstrate how to use our module for different tasks:

Runtimes on CPU Icosahesdron	**Order 3**	**Order 4**	**Order 5**	**Order 6**	**Order 7**
Augmentation	sec (↓)	sec (↓)	sec (↓)	sec (↓)	sec (↓)
SurfCutOut	10^{-4}	10^{-4}	10^{-3}	10^{-3}	10^{-2}
SurfNoise	10^{-5}	10^{-4}	10^{-4}	10^{-3}	10^{-3}
SurfBlur	10^{-4}	10^{-3}	10^{-2}	10^{-2}	10^{-2}
SurfRotation	10^{-1}	10^{0}	10^{0}	10^{1}	10^{2}
HemiMixUpHemi	10^{-5}	10^{-4}	10^{-4}	10^{-3}	10^{-3}
GroupMixUp	10^{-4}	10^{-3}	10^{-3}	10^{-3}	10^{-2}

Table 3.3: Execution time in seconds for each augmentation implemented on surfify. The augmentation were executed on a Intel(R) Core(TM) i7-3770 CPU, 3.40GHz.

- for prediction (classification or regression), we implemented various flavours of VGG [321] architectures.
- for segmentation (classification of every vertex in input space), we implemented the spherical U-Net [300, 394] architecture.
- for unsupervised representation learning and generative learning, we implemented a configurable VAE [197] architecture.

3.5.2 . Experiments related source code

All the code used to produce the experiments presented in Section 3.4 was released in an open source repository on github `https://github.com/neurospin-projects/2022_cambroise_surfaugment` and uses surfify, in which the proposed augmentations have been integrated.

3.6 . Conclusion

Modelling cortical surface as an icosahedral mesh is a very comprehensive way to incorporate the specific topology of the cortical surface of the cortex in the data. It accounts for the fact that the cortex is uninterrupted in healthy humans and the fact that to neighbour gyri are in fact biologically distant. Imposing such structure in data leads to many benefits. It can be used to reduce the dimensionality of the data, compared to raw 3D images, focusing on a signal of interest for studying psychiatric diseases, while being more precise and spatially coherent compared to average measures on ROIs. Additionally, having a data representation structured with biological considerations can allow subsequent modelling to take it into account and provide more informed and coherent biomarkers, thus improving interpretability.

In this context, we studied specific adaptations needed to build structure-aware cortical data representations using NNs. We leverage existing adapted operators such as DiNe, that allows us to design SCNNs able to learn from icosahedral cortical surface data. We introduce cortical

surface augmentations designed for training a SCNN in a self-supervised learning setup. The investigated SimCLR-SCNN shows the ability to generate representations with strong generalisation properties. In fact, the learned representations from data collected from multiple sites offer promising performance, sometimes even outperforming supervised approaches. In particular, the GroupMixUp augmentation shows potential for learning stable representations across different cohorts. Additional experiments could be lead to further investigate the characteristics required for cortical-based data augmentation. Moreover, an improvement to the GroupMixUp augmentation could incorporate prior information when computing the groups G_i, such as clinical scores. A similar strategy is proposed for structuring the learned representations by adding a regularisation term in the training loss [104].

Overall, we develop a methodology to learn generalisable representations of cortical surface data. It could be exploited to better initialise NNs on large healthy cohorts of population, before transferring to downstream clinical tasks of interest on smaller cohorts. We contribute to this higher purpose by offering a few cortical surface data augmentations, keystone to fit with novel self-supervised frameworks, and the surrounding tools required to use them.

The surfify module was presented during OHBM 2022 at Glasgow during poster sessions. Additionally, the experiments related to SSL on the cortical surface led to a publication in the Machine Learning in Clinical Neuroimaging (MLCN) workshop of MICCAI 2023. It answers the problem of using deep learning approaches from brain T1-weighted MRI while leveraging the cortical structure in this data. In the next part, instead of working on intra-modality structure, we will try to provide answers on how to use deep learning for modelling relationships between various sources of data in a population cohort, such as T1-weighted MRI with other imaging modalities, genotyping data or clinical questionnaires.

4 - Inter modality relationships modelling

Contents

In the previous chapter, we addressed the following question: how can we design DL models to leverage structure from data ? We used the eloquent example of representing the cortical structure derived from sMRI as a regular spherical mesh. In this context, we developed a SSL scheme, proposing cortical surface data augmentations and using specific convolution operators to maximally leverage the cortical structure to learn condensed representations. An interesting perspective of this work is to use this framework to initialise NNs on a large healthy cohort of population, in order to transfer them to smaller cohorts to extract clinically relevant information such as prognostic, segmentation or discovering biomarkers.
Others questions arise when one wants to exploit further available biological information: how can we take into account multiple data of varying nature, which are often assessed in population studies in a statistical model ? Then, how to extract knowledge from such a highly multivariate model ? Through this chapter we will try to answer them to some extent, with an application in psychiatry, where we believe integration using Probabilistic Graphical Models (PGMs) are of particular interest.

4.1 . Problem statement

The hope of integrating multiple views representing various types of data is to produce richer models that better account for the complex context of a measurement on a subject. Such models are expected to fit better the data, taking into account multi-scale biological processes, and thus produce more consistent findings. Another hope is to produce explainable models that would leverage the different modalities of the measurements. Explanations can be presented as models for prediction to build decision support systems, or to discover biomarkers, which in turn can help in understanding biological mechanisms underlying the phenotype. The ultimate goal of such methodology is to find treatments to related diseases.

This integration problem is particularly expressed and experimented when studying brain pathologies such as cancers or psychiatric diseases. These pathologies are mostly not understood today, probably because the brain is the most complicated human organ. However we can have access to a multitude of biological assessments thanks to the developments of high throughput biological tools such as genotyping, MRI for brain structure and histology for tissue inspection, that we hope can help us understand brain disorders. Integration opposes to classical massively univariate approaches such as Genome Wide Association Study (GWAS) for genotyping data, Statistical Parametric Map (SPM) for functional MRI, or approaches considering only one modality, which is commonly done in diagnostic predicting models.

Heterogeneity in psychiatric diseases. For many years, researchers have tried to find biological correlates of diseases using mostly one biological assessment (e.g. sMRI or genotyping) with respect to a binary or categorical diagnostic. For brain disorders, these diagnoses are rigorously defined in the successive versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM) [369, 246, 78]. However, results highlighted throughout these studies have shown to be rarely reproducible [22, 269, 381], and even then not necessarily understood.

The major hypothesis arising from this observation is that mental disorders are highly heterogeneous [114, 7]. This implies that trying to find biological markers of pathology relying solely on DSM diagnostics using one biological assessment is probably not efficient to discover the complex mechanism hiding behind behavioral symptoms, which can be common to multiple diseases. The Research Domain Criteria (RDoC) [181] advocates for a more comprehensive approach to studying psychiatric disorders by incorporating diverse data types that cover different levels of life organisation (e.g., imaging, genetics, symptoms). The RDoC principles suggest that a thorough description of a pathology requires consideration of multiple "latents" (also called "dimensions") that may be shared across different psychiatric syndromes and may even contribute to non-pathological variability. Notice that the term "latent" (or "dimension") must be considered here in the context of the National Institute of Mental Health (NIMH) RDoC approach which is one framework among others that study psychiatry.
The approaches that take into account such recommendations include two major identified questions, which created opposing but not necessarily mutually exclusive [382, 253] lines of research.

The first one investigates if we can find homogeneous subgroups within pathological patients that present similarities, either at the phenotype level or at the biological level, in order to design targeted treatments for individuals within specific subgroups or subtypes, also called population stratification [310, 114, 17]. These approaches usually rely on a clustering algorithm. However, this subgroup discovery approach has been criticised, and only a few studies have produced consistent and reproducible groups across studies [111, 243]. Moreover, they usually focus on a particular population related to a single diagnosis, ignoring the recommendations of the RDoC, even if some studies propose transdiagnostic stratifications [148, 280, 331].

The second line preaches for a dimensional approach, which means discovering low dimensional manifolds expressing variability somehow related to expressed pathological symptoms. Studies investigating brain disorders this way prefer considering patients on a spectrum with respect to their symptoms, spectrum on which healthy individuals are represented as well [242, 241].

Availability of multimodal data in psychiatric population studies. In line with the RDoC principles, psychiatric research has made a huge effort for building large cohorts to study psychiatric syndromes. We can distinguish various types of cohorts that aim at answering different research questions.

Depending on the studied disorder, psychiatric cohorts recruit subjects of varying age [185, 324]. For developmental disorders, cohorts often recruit from early childhood ($\sim$ 4 years old) to early adulthood ($\sim$ 25 years old). Among them, most studies are based on a case-control recruitment, because they are built to investigate particular pathologies, such as Autism Spectrum Disorder (ASD) [98, 99, 263] or Attention Deficit Hyperactivity Disorder (ADHD) [256]. Other studies recruit either using very few criteria, or simply targeting an at-risk population for multiple disorders, meaning the children present some psychiatric symptoms without necessarily being diagnosed. These cohorts can be designated as transdiagnostic cohorts and are used to discover markers of abnormal

development [52, 6] across multiple disorders. There exist adult cohorts aswell, which can be either case-control [340, 360] or transdiagnostic [284], designed to study adulthood markers of psychiatric disorders.

The statistical power of large cohorts offer promising perspectives to find relevant biomarkers for psychiatric disorders [240]. However, they are often impaired by the fact that they are multi-centric, which brings differences in scanning system, settings, acquisition conditions and prevalence specificities. This site effect has been extensively described and how to handle it is still an open research question [105, 175]. These studies can have single or multiple assessments per individual, which we refer in the latter case to multimodal studies. These assessments can be evaluated one time for each individual or at multiple time points in longitudinal studies. The longitudinal setting allow researchers to investigate personalised atypical trajectories, but they are very expensive and long to build, and often present at lot of missing data. Usual biological proxies assessed in these studies include (but are not restricted to) imaging through structural MRI, resting-state functional MRI, task-oriented functional MRI, diffusion MRI, EEG and MEG, and genotyping as SNP and CNV data. Such cohorts often include phenotype data, which regroup morphological assessments, behavioural data represented as a diagnosis or clinical questionnaires evaluating symptoms, and environmental factors. Finally, these studies crucially also provide access to demographics consisting in the age, sex and imaging acquisition centre, which often have interactions with diagnoses.

In order to handle the growing quantity and variety of data and model their relationships, with respect to disease related variability, we need machine learning tools with the following properties:

- **Unsupervised or semi-supervised**: the learning algorithm should not be only directed by a diagnostic prediction, which does not express sufficient pathology related variability, and limits the quantity of available data. Considering a semi-supervised setting instead, by integrating symptoms instead of diagnosis, offers to consider both subjects with symptom related assessments and subjects exempt of them.
- **Integrative**: able to handle an arbitrary amount of data in various forms, and model their shared variability, while not over-sighting intra-modality correlation structure.
- **Able to handle missing data**: all multimodal cohorts have missing data, that can appear at different levels, either sporadically in one type of assessments, such as a non fulfilled clinical questionnaire, or a completely missing modality such as sMRI, if the patient could not be scanned for some reason. Such missing occurences are very frequent in multimodal settings, and the ML model we choose should be able to use observations with missing data, otherwise risking of being statistically greatly impaired.
- **Interpretable**: we should be able to understand and explain what the model has learned. The ultimate goal would be to outline biological mechanisms responsible for psychiatric diseases, discovered by the model. This is impossible if the model is not somehow interpretable.

4.2 . Existing multimodal approaches and their properties

In this section, we want to sketch up a state-of-the-art of the existing approaches matching some of the characteristics describes above. Multimodal integration were first developped to answer to specific applications. Depending on the problem to solve, researchers have developed an enormous quantity of methods that can be qualified as multimodal. Multiple taxonomies and reviews [25, 128, 29] have been proposed. We are not aiming to describe all existing methods and learning schemes, but rather describe the few methods that answer our requirements explicited in the previous section, some of their history, properties and applications. In particular, we start by describing some linear integration models, a few extensions and some applications in the medical field. Then we introduce some relevant NN based methods, namely multimodal PGMs. We describe most of their extensions and a few applications in healthcare.

4.2.1 . Linear integration models

Pioneering works on integrative models are Canonical Correlation Analysis (CCA) [174] or Partial Least Squares (PLS) [371]. We retain these two methods as emblematic of integration analysis and describe them further below. They can be identified as latent variable models, because they are based on the description of unidimensional latent variables, which are a linear projections of the variable in input space. We start by describing their formulations, some of their extensions and a few applications.

Let $x = \{x_1, \dots, x_k, \dots, x_K\}$ be an observation of K views, where each x_k can have different dimensions J_k. Let's assume we have access to n such observations $\mathbf{x} = \{x^1, \dots, x^n\}$, for which all K views are available. Let's regroup these observations view wise, noting $\mathbf{x}_k = [x_k^1, \dots, x_k^n]^\intercal \in \mathbb{R}^{n \times J_k}$, for $k \in \{1, \dots, K\}$. The objective of these model is to find weights $w_k \in \mathbb{R}^{J_k}$ such that the projections $\mathbf{x}_k w_k$ have maximal correlation (CCA) or covariance (PLS) with each other. More formally, CCA finds the solution to

$$\underset{\{w_1,\dots,w_K\}}{\text{argmax}} \sum_{k=1}^{K} \sum_{j=k+1}^{K} c_{kj} \text{corr}(\mathbf{x}_k w_k, \mathbf{x}_j w_j)$$

where PLS searches for the solution to

$$\underset{\{w_1,\dots,w_K\}}{\text{argmax}} \sum_{k=1}^{K} \sum_{j=k+1}^{K} c_{kj} \text{cov}(\mathbf{x}_k w_k, \mathbf{x}_j w_j) = \sum_{k=1}^{K} \sum_{j=k+1}^{K} c_{kj} \text{corr}(\mathbf{x}_k w_k, \mathbf{x}_j w_j) \sqrt{\text{var}(\mathbf{x}_k w_k)} \sqrt{\text{var}(\mathbf{x}_j w_j)}$$

where $C = \{c_{kj}\} \in \{0, 1\}^{K \times K}$ is the triangular superior design matrix, which defines which views are connected in the model and is set by the user. These optimisation problems can be solved in different ways, using their eigen value decomposition [174] or Nonlinear Iterative Partial Least Squares (NIPALS) [344, 303] for instance.

Multiple extensions have been proposed, notably Generalised CCA (GCCA) [50] , Kernel CCA (KCCA) [77], probabilistic CCA [21], and bayesian CCA [202, 358], allowing more flexibility to the framework.
On the other hand, CCA and PLS have been shown to work poorly in particular settings, especially when the number of features J_k within a view is higher than the number of observations n or when

features within a view are highly correlated [356]. The most frequent solutions to this problem are using a PCA to reduce dimensionality of concerned views, or penalise the optimisation problem using classical regularisations such as l_1 or l_2 regularisations [303, 154, 343], therefore constraining the learned weights. It is interesting to note that the l_2 regularisation in particular unifies both approaches, as PLS can be viewed as a maximally l_2 regularised CCA [303, 154, 343]. Additionally, l_1 regularisation [370, 342] can be viewed as a sparsity constraint, in other words as a variable selection regularisation. Both regularisations are often used jointly, implementing the elastic net regularisation [395], which combines their properties. Finally, another GraphNet penalty [224, 149] has been proposed for adding expert knowledge in the learning algorithm.

These integrations methods have been successfully applied to multiple fields [25]. In health applications, it has been recurrently shown that integrating complementary modalities leads to increased performance for the task of interest [204, 240, 329, 177]. Most common applications include oncology and neurology [204], where multiple assessments can contribute jointly.
In oncology, linear integrative models have been used for survival prediction using highly multimodal cohorts such as TCGA [365, 229]. Such cohorts allow integrative model to exploit multi-scale information from both imaging and multi-omic modalities acquired from biopsies to predict survival with the Cox model [76]. Linear integration tools such as CCA or PLS have been largely benchmarked for this survival prediction task [163, 290]. Interestingly, adding biological a priori to the learning algorithm using protein-protein interaction networks with GraphNet has been shown to improve the model survival predicting performance and variable selection [150, 57].
In neurology, linear integration has often been applied to find links between an observed phenotype and an imaging modality [260, 135], or between different biological assessments, such as SNP data with fMRI [216].

However, these methods still have some limitations:

- CCA maximises correlation between latent variables from each view, so it is more sensitive to the direction of the relationships between modalities, and is not impacted by within modality variance. PLS is driven by covariance, so it is less directed by inter-modality correlation structure, but is sensitive to intra-view variance [254].

- Additionally, most of these linear methods have a simplistic way of taking into account observations with missing data. They often rely on imputation techniques such as the mean feature value or 0. However, this strategy can introduce issues and many applications discard observations with missing values. In multimodal cohorts, complete views are often missing, as they represent a given biological assessment. The number of observations with all modality inevitably decreases as the number of assessments grows. This is why it is important for methods to be able to take into account observations with missing views.

- Moreover, most CCA and PLS derived algorithms are linearly constrained, and even their non-linear extensions such as KCCA, remain computationally expensive in high dimension, still approximate a restricted class of function by the kernel choice and remain constrained by the properties of CCA, such as the fact that their are poorly effective at modelling within-modality

structure. An extension using NNs to learn the latent variable was proposed [13] to overcome some of these limitations, paving the way for NN based integration models.

4.2.2 . Deep learning based approaches: probabilistic graphical model.

Multimodal integration tools are always developed at first to answer specific applications. Of particular interest are the PGM based integration methods. They naturally embed interesting properties, such as missing data robustness [25], scaling to an arbitrary number of modalities (for some of them at least) and display very relevant disentanglement capabilities.

Early NN based multimodal integration tools using PGMs [327, 266] were proposed as extensions of Restricted Boltzmann Machine (RBM) [165]. A multimodal Deep Boltzmann Machine (DBM) [328] architecture was further introduced to integrate an arbitrary number of modalities. However, these techniques are computationally heavy, as they use Monte Carlo Markov Chain (MCMC) sampling, which makes them hard to scale to high dimensional inputs. Worth noting as well, a non-linear extension of CCA called Deep CCA [13] was proposed as an alternative to kernel CCA, which is more efficient and offer a wider class of function that can be learned using NNs. It was then further extended to the reconstruction based framework Deep Canonically Correlated Auto-Encoders[361], bridging the gap between CCA and auto-encoders.

Kingma et al. [197] proposed a learning scheme to do variational inference using NNs. This groundbreaking work led the PGM multimodal community to shift from MCMC based inference to using VAEs as it is more efficient to approximate distribution parameters. In particular, it introduced Variational Auto-Encoder (VAE) as a regularised auto-encoder, using the divergence of Kullback-Liebler between the latent space posterior distribution and a Gaussian prior for regularisation. There have been multiple multimodal extensions proposed from this initial work [337]. The same authors proposed a semi-supervised extension [196] of VAEs, allowing to condition the latent space on a label when available. It can be seen as an asymmetrical multimodal integration technique. Other asymmetric multimodal PGMs where proposed [275, 187], but it was then shown that symmetric conditioning have better generative properties [338]. In another line of work, Deep Variational CCA [362] was proposed as an extension of probabilistic CCA. It interestingly introduces the concept of private latent spaces, showing that it improves reconstructions and disentangles the information contained in the different latent representations.

In the next section, we introduce a particular type of PGMs which relies on VAEs, that we define as multi-view Variational Auto-Encoder (mVAE), with usual assumptions about there generative and inference models. We then describe two particular mVAEs with interesting properties, and introduce a few of their applications.

Multi-view VAEs. Let us introduce more formally the multi-view Variational Auto-Encoder (mVAE) learning framework to further clarify. Let $x = \{x_1, \ldots, x_k, \ldots, x_K\}$ be an observation of K views, where each x_k can have different dimensions J_k. We assume that their exist $K+1$ multidimensional latent variables underlying the data generation process, one latent variable specific to each view, and

(a) Only joint. (b) Joint and specific.

Figure 4.1: Two different PGMs illustrating the generating process modelled in various mVAEs in case of two modalities. (a) Joint model such as in [372, 318, 336]. (b) Explicitly modelling modality-specific and shared latent variables as proposed by [362, 348, 222].

one joint, expressing the shared variability across views (see Fig. 4.1). The general form of these latent variables is $\mathbf{z} = \{z, z_1, \ldots, z_k, \ldots, z_K\}$ with z denoting the d-dimensional latent variable shared by all x_k, and z_k the view-specific d_k-dimensional latent variables. We assume the following generative process for the observation set:

$$\begin{aligned} z &\sim p(z) \\ z_k &\sim p(z_k) \\ x_k &\sim p_\theta(x_k|z, z_k) \end{aligned}$$

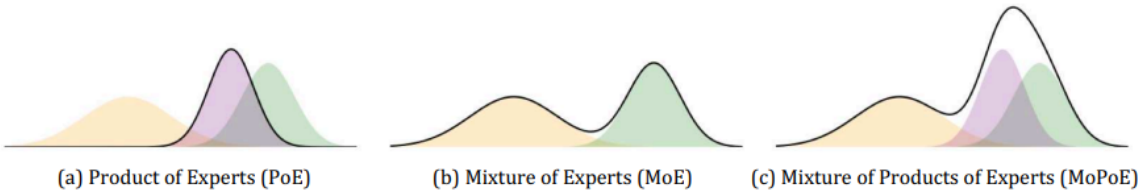
where $p(z), p(z_k)$ are prior distributions for the latent variables and $p_\theta(x_k|z, z_k)$ is a likelihood distribution of the observations conditioned on the latent variable. The bottleneck of maximising this likelihood comes to computing the true posterior $p_\theta(\mathbf{z}|x)$. A first usual assumption about the true joint posterior $p_\theta(z|x)$ considers that it contains all the shared information between the views, and we can write $p_\theta(\mathbf{z}|x) = p_\theta(z|x) \times \prod_k p_\theta(z_k|x_k)$. Other assumptions are needed since these posteriors are not analytically tractable. We defined the approximated variational posterior $q_\phi(\mathbf{z}|x)$. The variational parameters are described by ϕ where θ describes the generative parameters. In the same way, we have $q_\phi(\mathbf{z}|x) = q_\phi(z|x) \times \prod_k q_\phi(z_k|x_k)$ These posterior distributions are classically defined as parameterised distributions $q_\phi(z|x)$ (the so-called joint variational posterior) and $q_\phi(z_k|x_k)$ respectively. The generative process $p_\theta(x)$ is then approximated by its Evidence Lower Bound (ELBO):

$$\begin{aligned} \log p_\theta(x) &= \log \int p(x, \mathbf{z}) \frac{q_\phi(\mathbf{z}|x)}{q_\phi(\mathbf{z}|x)} d\mathbf{z} = \log \mathbb{E}_{q(\mathbf{z}|x)} \left[\frac{p(x, \mathbf{z})}{q_\phi(\mathbf{z}|x)} \right] \\ &\underbrace{\geq}_{\text{Jensen}} \mathbb{E}_{q_\phi(\mathbf{z}|x)} \left[\log \frac{p(x, \mathbf{z})}{q_\phi(\mathbf{z}|x)} \right] = \mathbb{E}_{q_\phi(\mathbf{z}|x)} \left[\log \frac{p_\theta(\mathbf{z}) p_\theta(x|\mathbf{z})}{q_\phi(\mathbf{z}|x)} \right] \\ &\geq \mathcal{L}_{\text{ELBO}}(x; \theta) \end{aligned} \tag{4.1}$$

The ELBO loss $\mathcal{L}_{\text{ELBO}}(x; \theta)$ can be written as

$$\begin{aligned} \mathcal{L}_{\text{ELBO}}(x; \theta) &= \mathbb{E}_{q_\phi(\mathbf{z}|x)}[\log p_\theta(x|\mathbf{z})] + \mathbb{E}_{q_\phi(\mathbf{z}|x)} \left[\log \frac{p_\theta(\mathbf{z})}{q_\phi(\mathbf{z}|x)} \right] \\ &= \mathbb{E}_{q_\phi(\mathbf{z}|x)}[\log p_\theta(x|\mathbf{z})] - D_{\text{KL}}(q_\phi(\mathbf{z}|x) \| p_\theta(\mathbf{z})) \end{aligned} \tag{4.2}$$

State-of-the-art models solve this problem with different ways of modelling the joint variational posterior distribution $q_\phi(z|x)$. These approaches compose $q_\phi(z|x)$ using the Gaussian experts $q_\phi(z|x_k)$ unimodal posteriors, which may be combined in various ways. In Product of Experts (PoE) approach [372], $q_\phi(z|x)$ is defined as a product of individual Gaussian experts, assuming the conditional independence of the experts, hence remains Gaussian. The product operator used in PoE may dampen the specific variability of a given modality described by Gaussian unimodal posterior, and other work proposes to model [318] $q_\phi(z|x)$ using a Mixture of Experts (MoE), assuming each unimodal expert contributes equally to the joint posterior. MoPoE approaches defines $q_\phi(z|x)$ as a Mixture of Product of Experts. Each product of experts is computed over the different subsets of views [336]. These different forms of joint posterior distribution are illustrated by the black line in Figure 4.2 with a 1-dimensional example. The MoPoE-VAE is a generalisation of the two previous approaches and combines their advantages, i.e. properly models the shared variability between modalities while considering their individual contribution. Each strategy has different ways of handling missing views, that may differ at train and test time. The MoPoE-VAE simply uses all the subsets of available views for each observation. As regards to the unimodal posteriors $q_\phi(z_k|x_k)$, we classically use a Gaussian distribution.



(a) Product of Experts (PoE) (b) Mixture of Experts (MoE) (c) Mixture of Products of Experts (MoPoE)

Figure 4.2: Illustration of the existing joint posterior models from [337].

The MoPoE-VAE with view-specific and shared latent spaces. The MoPoE-VAE optimises a generalised multimodal ELBO objective for learning view-specific and a joint distribution of multiple views x with potential missing data. This model assumes conditional independence between the marginal posterior distributions, i.e. $p_\theta(x|\mathbf{z}) = \prod_k p_\theta(x_k|\mathbf{z})$, and simplifies the marginal posteriors $p_\theta(x_k|\mathbf{z}) = p_\theta(x_k|z, z_k)$ using the conditional independence between x_k and z_j given z, z_k with $j \neq k$. The mulitmodal ELBO objective is the following:

$$\mathcal{L}_{\text{MoPoE}}(\theta, \phi; x) = \sum_{k=1}^{K} E_{q_\phi(z,z_k|x)} \left[\log(p_\theta(x_k|z, z_k)\right]$$
$$- \sum_{k=1}^{K} D_{KL}\left(q_\phi(z_k|x_k)||p_\theta(z_k)\right) - D_{KL}\left(\underbrace{\frac{1}{2^K}\sum_{x_p \in \mathcal{P}(x)} \tilde{q}_\phi(z|x_p)}_{=q_\phi(z|x)} ||p_\theta(z)\right)$$

There are 2^K different subsets contained in the powerset $\mathcal{P}(x)$, and $\tilde{q}_\phi(z|x_p) = \prod_{x_k \in x_p} q_\phi(z|x_k)$ is the product of Gaussian expert posteriors $q_\phi(z|x_k)$, defined for the subset of views in x_p. This gives the ability to the MoPoE-VAE to handle any subset of view at training and inference time, in exchange with an elevated computational cost, scaling with the number of modalities.

These formulas illustrate a specific implementation of the MoPoE-VAE which handles multi-views by modelling view-specific latent space posterior distributions, and a joint posterior distribution shared between views as a mixture of the products of their marginal posteriors (experts). This implementation differs from the one benchmarked in [88, 337, 312] which consider only one joint latent (MoPoE-VAE with its initial formulation [336]).

The Multi-Channel VAE and its sparsity constraint. The MCVAE [14] uses slightly different assumptions as the ones presented above. It does use the conditional independence assumption $p_\theta(x|z_1, \dots, z_K) = \prod_k p_\theta(x_k|z_1, \dots, z_K)$. Contrary to the MoPoE-VAE, it can not hypothesise for conditional independence between x_k and z_j given z_k when $k \neq j$ because it does not explicitly model a joint latent space, otherwise it would result in K independent parallel VAEs. These assumptions result in the following loss:

$$\mathcal{L}_{\text{MCVAE}}(\theta, \phi; x) = \frac{1}{K} \sum_{k=1}^{K} \left(\mathbb{E}_{q_\phi(z_k|x_k)} \left[\sum_{c=1}^{K} \log p_\theta(x_c|z_k) \right] - D_{\text{KL}}(q_\phi(z_k|x_k)|p(z_k) \right)$$

Another significant contribution of this work is to extend an existing sparsity constraint to a multi-view setting. This sparsity constraint amounts to hypothesising a particular form of the marginal approximated variational posteriors $q_\phi(z_k|x_k) = \mathcal{N}(\mu_k, \alpha\mu_k^2)$ with $\alpha \in \mathbb{R}^{d_k}$. However this assumption requires using slightly different priors, namely the scale-invariant log-uniform priors, i.e. such that $\log p(|z_k|) \propto$ const.. In this case, an approximation of the $D_{\text{KL}}(q_\phi(z_k|x_k|p(z_k)))$ has been proposed in [259], which depends only on α. This allows α to be approximated using gradient descent. Having the same latent dimension for all the latent variables is a requirement in the MCVAE in order to easily provide each latent representation to the decoders during training. They propose to optimise the MCVAE objective, but with these log-scale uniform priors and these particular approximated posterior form, and interestingly to use the same parameter α across every modality's latent space. As this parameter is commonly optimised during training for all latent space, we can compute a threshold over the dropout rate $\frac{\alpha}{\alpha+1}$ for each latent space dimension, above which the dimension corresponding weights in the different encoders can be dropped (set to 0). It thus forces the latent spaces to have the same dimensions, and regularises all the latent spaces with a shared constraint. However, this initial formulation does not allow to use data examples with missing views. It was later extended to adapt the learning algorithm to data with missing views [15].

Most recent works focus on modelling, additionally to the joint latent space, a private (or specific) latent space for each modality [348, 335, 222, 274, 285, 188], demonstrating improved generative abilities and disentangling of modality-specific and shared factors. However, it was shown that without explicitly enforcing disentanglement by constraining the learning for instance, fully disentangling

these factors is not possible [232]. Some works have proposed techniques to improve the disentanglement of view-specific and shared factors for these SotA multimodal PGMs [317, 375, 191, 89].

Applications of mVAEs. Multimodal machine learning tools have a large panel of applications including audio visual speech recognition applications, multimedia retrieval and media description or generation [25], such as Visual Question Answering (VQA) [16]. We will focus on applications in the medical field, as they illustrate the variety of tasks that can be apprehended using integration models, being one of the application domain most explored by these multimodal techniques [29]. In the medical field, similarly to other multimodal integration models [204], frequent applications of multimodal PGMs include oncology and neurology.

When referring to multimodal integration with deep PGMs in oncology, we can identify multi-omic integration applications. An information bottleneck approach [221] uses the variational approximation of the joint posterior as a PoE and view-specific latent representations from multi-omic data. The joint representation is further supervised by the task of interest, such as survival prediction on the TCGA dataset. Another VAE-based approach [388] concatenates each modality latent representations rather than explicitly modelling their joint posterior distribution to predict tumour type from multi-omic data aggregated from TCGA pan-cancer datasets.

In neurology, a very frequent application of multimodal machine learning is Alzheimer's disease (AD) prediction [204, 258]. In fact, AD prediction was a task addressed very early using deep PGM integration model [333] exploiting PET and MRI, and later to showcase the sMCVAE capacities [14]. Also worth mentioning, the same way multimodal PGMs shifted from using MCMC sampling for doing variational inference to using VAEs, Normative Modelling (NM) [242] using VAEs have been proposed [214, 363]. They offer the opportunity for NM to scale to larger cohorts and to be multivariate. Pushing even further the multivariate paradigm, some multimodal normative VAEs have been proposed [209, 208, 213], leveraging recent advances of mVAE and applied to AD study. Yet, they have not fully seized the opportunities offered by mVAEs, as none of them model modality-specific latent spaces.

Multimodal integration has not yet reached its full power, but it is already widely exploited and developed. We identified mVAEs as good candidates to answer the problem introduced in the previous section, presenting most of the characteristics we described earlier. They are able to integrate an arbitrary number of modality and take modality structure into account by using modality-specific encoders and decoders. Handling missing modalities naturally arises from their variational aspect, although it is handle differently, depending on the variational approximation of the joint latent posterior. When diagnostic or clinically relevant symptom severity scores are only available for a part of the participants, it thus creates a weakly-supervised scheme because mVAEs are able to learn from participant data with missing modalities. Beyond their suitability for multimodal integration, VAEs are known for their interpretability capacities. In the following paragraph, we introduce the interpretation capabilities of the VAE in the context of multimodal integration models. We specifically present the instability sources to which these models (and thus their interpretations) are subject to, which should be properly addressed to produce reproducible findings.

4.2.3 . Interpretation

Interpretation of multimodal NNs is usually focused on decision aiding tools and supervised models [186, 294]. Most of these methods have been developed for image or text explanations, specific to the NN architectures and learning tasks. In particular, the wide majority of these methods have been developed for VQA applications. In a comparable way, we aim to assess what our mVAE has learned in human readable terms. For instance, linear associations are very eloquent and are trivially available for linear integration methods such as CCA or PLS. Their l_1 regularisation additionally offers a feature selection mechanism that provides even more readable interpretations thanks to the sparsity of the solutions. This strategy is not directly applicable in DL, where the NNs often learn complex non-linear functions.
Additionally, ensuring that findings (e.g. interpretations) are stable to variation in experimental setting is crucial to allow for reproducibility and turn results into knowledge. There are various sources of variability that can impact the outcomes of the integration models we introduced, but they can be qualified and properly handled. In the following, we first discuss interpretation opportunities in VAEs, and then describe more broadly sources of instability in machine learning, which are both relevant to discuss mVAE interpretation.

Usual methods for interpreting VAEs and their issues

VAEs are unsupervised models, that can be used either for their capacities to learn representations, or for their generative properties. Interpretation in VAEs stems from their generative characteristics and surprising disentanglement capabilities.

Interpretation from browsing the latent space. A usual way to assess what a VAE learned is to sample in its latent space, then reconstruct the corresponding data point using the decoder. Usually, instance-level or group-level explanation levels are generated by either considering two instances latent representations, sampling in between using interpolation, or average latent representation from the latent representations of a group of instances. Their reconstruction should evolve in a smooth way and describe properties of the latent space. This is due to the fact the VAEs learn continuous functions, which comes from the variational regularisation of the latent posterior distribution to the same prior for all observations [197]. The same procedure can be done the other way round: modifying smoothly the input data and observe the changes in latent representation space can help investigate how the latent space is structured. For small dimensional (2D or 3D) latent space, one can sample regularly in the latent space and reconstruct them to characterise the whole space by mapping it with characteristics from the input space. Most often, latent spaces have a greater dimension and this is not applicable, and an additional dimension reduction technique (e.g. PCA or t-SNE) must be used to offer readable visualisations.

Variability separation by constraining the latent space with β. In 2017, a paper [164] shows that increasing the regularising parameter β of the VAE, which penalises the latent posterior

distribution, provides disentanglement, in that human understandable concepts are separated in different dimensions. For instance, the dataset dSprites[1] consists of black images with white shapes embedded in each image. These shapes are encoded by configurable parameters of shape (square, ellispe, heart), position, rotation and scale. They showed that increasing β leads to isolate these parameters in different latent dimensions. Many works have tried to investigate this effect [45, 59, 244]. However, discussing such disentanglement properties is debatable, as it was shown that such effect might not be necessarily due to proper disentanglement. Indeed, it was proven impossible to disentangle sources of variability in a fully unsupervised setting without inductive bias or other imposed constraints [232].

Although these approaches allow to interpret VAEs latent space, the provided interpretations are instance or group wise. Moreover, relying on the latent space to explain learned associations by the model seems to be difficult, because we are unable to fully understand and hence control this space. In the next section, we will introduce the Digital Avatar Analysis (DAA), which offers feature wise inter-view interpretations of the joint relationships learned by a multi-view Variational Auto-Encoder. It relies on linking together reliable modifications in input X-space to output (also X-)space.

Stability issues in mVAE interpretability

Statistical learning models can be affected by numerous sources of variations, which can lead to varying modelling results. Some works identified most of these sources [201], and classified them into two families: aleatoric uncertainty, which is due to characteristics of the data and is thus irreducible from a modelling perspective, and epistemic referring to the reducible part of uncertainty coming from modelling choices. These concepts have been studied essentially in prediction settings and they are usually solved using Bayesian inference [179].

In parallel and in a more general context than mVAE interpretation, some works have defined the notion of stability [250], which refers to the stability of learned weights, under different hyper-parameter settings and across different training subsets of a population. The stability is often used to assess reproducibility of machine learning models, and in particular in the linear integration community with sparsity regularisation [342].
In this section, we discuss how these sources of variations are taken into account in these different domains, by first discussing linear integration approaches to assess this stability and their conclusions, then we introduce a few concepts aiming at reducing epistemic variability in NNs.

Stability of linear integration - handling aleatoric variability. In linear integration models, the main source of variability remain aleatoric. Indeed, the learning algorithm being constrained to linearity usually converges to the one solution, provided a given training population. Its epistemic variability amounts to the hyper-parameter choice, such as regularising constraints strength. Some of these integrative methods rely on sparsity constraint to identify the most relevant features,

[1] https://github.com/google-deepmind/dsprites-dataset

while other focus on features with the weights of largest amplitudes. Classically, these integration studies employ cross-validation strategies that often include the definition of a downstream auxiliary predictor for model selection [260]. However, it has been shown that these procedures may not consistently identify stable associations [215, 342], especially when correlations between views are low or sample sizes are small [215, 161, 376, 264]. To overcome this stability issue, some have proposed to use stability as a criterion in their hyper-parameter or variable selection process [152, 210, 23, 180, 255, 57], drawing inspiration from stability selection [250].

Sources of variation in Neural networks - handling epistemic variability. Over-parameterised neural networks have been shown to be far too confident about their predictions [151], which is a bad property especially when they are designed to be part of a critical application such as decision aiding tools. NNs uncertainty needs to be properly addressed, and it is related to the surprising generalisation properties of such over-parameterised networks (see Section 2.2.1). Most of the approaches that address this question consider Bayesian NNs. These particular NNs feature SotA uncertainty modelling properties, although they usually account exclusively for aleatoric or epistemic uncertainty [126, 192, 132, 179]. Bayesian NNs differ from deterministic NNs in that they put a prior on the weights of the model. Instead of optimising directly the weight values, they optimise the parameters of their associated prior distributions. For instance, a VAE can be seen as a special case of Bayesian NN where only the weights before the latent space are stochastic. It is worth mentioning the Monte-Carlo Dropout [127] that models epistemic uncertainty by introducing a Bernouilli distribution prior on the weights of a NN. This work explains quite elegantly the boosted generalisation properties observed when regularising a network with Dropout [326], which consist in randomly dropping weights (i.e. setting them to zero) during training. Additional work propose to model epistemic variability using ensembling [211], which consist in training multiple NNs over the same training set, simply varying the random weight initialisation and training batches. This frameworks conveniently extends epistemic uncertainty modelling to all types of NN trained in a supervised manner.

With this in mind, we can properly identify sources of variability that we should take into account in our mVAE interpretation framework: epistemic uncertainty will mostly arise from NNs training, given a fixed architecture, while aleatoric uncertainty will come from the training / left-out split for NN training and interpreting. In the next sections, we use these observations to derive an interpretation pipeline, with different building blocks, which when combined together should be able to extract inter-modality associations that are stable across these various sources of variability.

4.3 . Digital Avatars Analysis: leveraging generative abilities for interpretability

Interpretations in input data (X-)space can be achieved using Digital Avatars (DAs) by leveraging generative capacities of VAEs. Let us consider a trained mVAE. We consider L observations (or subjects) $\{x^1, \dots, x^L\}$ forming the left-out set, that our trained mVAE did not use for training. The L observations have all views available, i.e. $x^i = \{x^i_1, \dots, x^i_K\}$ for all $i \in \{1, \dots, L\}$ with K being the

number of views (e.g. imaging, questionnaires, genetic). Once trained, a mVAE cannot directly provide inter-view associations. However, the information learned by the model can be used to discover relevant pairwise associations of one view with the others. Let $l \in \{1, \ldots, K\}$ be this view's index. Starting from the left-out subjects, we modify each subject's features of this view l and generate a set of $T = 200$ DAs. Thus, as being predicted by the model, we capture the influence of this feature on other views within the model. From this set of virtual data, a conventional association study is performed to find inter-view relationships. Hereafter, we describe in detail how the DAs are obtained and analysed.

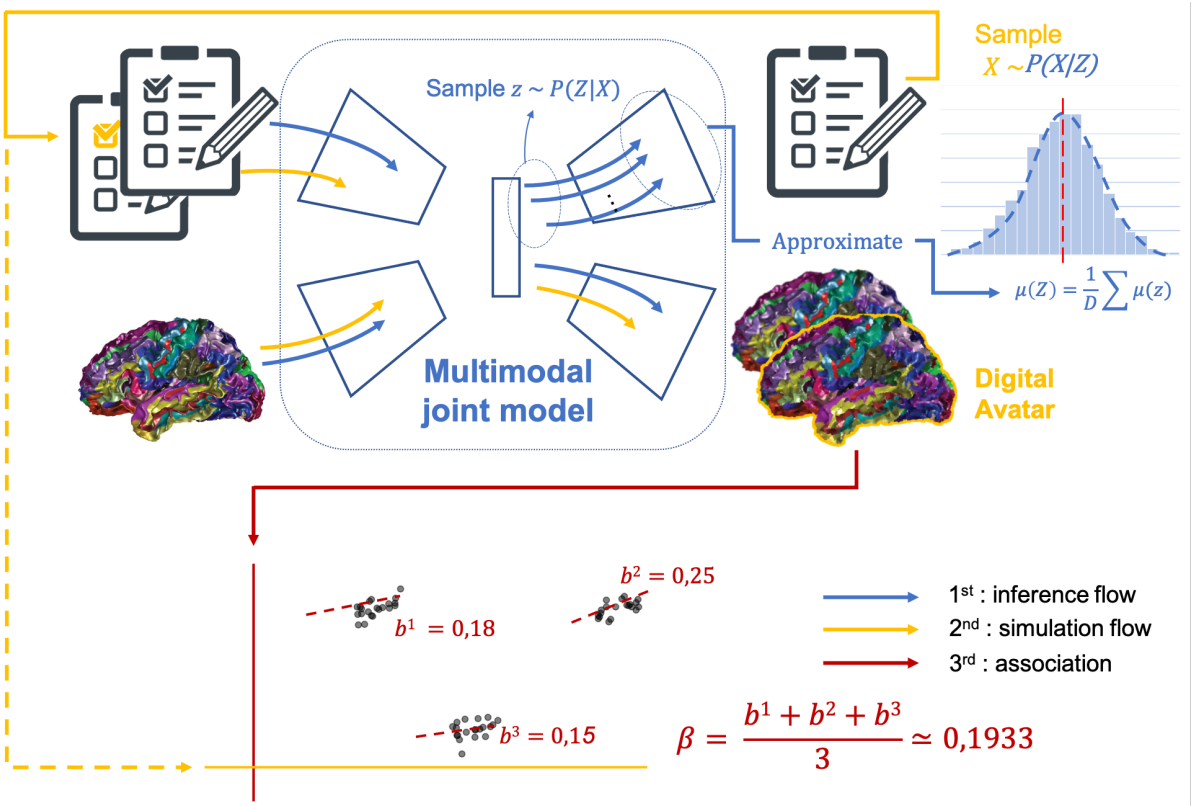


Figure 4.3: Illustration of the DAA interpretation framework in a clinical cohort setting with two modalities: imaging data and clinical questionnaires. First, the inference flow estimates output distributions via sampling in the latent space. Then, the simulation flow generates realistic perturbed samples of the view we want to study against others (here, the questionnaires) and infer digital avatars through the model. Finally, meaningful inter-view associations are inspected using hierarchical linear regressions.

Without loss of generality and for simplicity, we consider the case where there are two views, the index of the reference view $l = 1$ and we note $x^i = (s^i, m^i)$. To further simplify notations, we assume view l only has one feature. Our aim is to create a set of DAs from each of the L left-out subjects. Let $s = (s^1, \ldots, s^L) \in \mathbb{R}^L$ represent the view l, and $m = [m^1, \ldots, m^L]^\intercal \in \mathbb{R}^{L \times p_m}$ the other view, p_m being the number of feature of this view. Note that this specific case can be generalised to cases where view l has multiple features and additional views. The general idea is to modify s into $\hat{s}$ and observe how

these modifications affect the reconstructed features $\hat{m}$. In this way, virtual pairs ($\hat{s}$, $\hat{m}$) are obtained. A first, although simplistic, approach to perturb s is to sample linearly or randomly between feature percentiles or even to bootstrap values across subjects. Conversely, the proposed approach relies on a simulation scheme to realistically perturb s. The realism of these perturbations stems from the generative aspect of the model. We hypothesise that such an approach, which models the composite variability of subjects measurements, will facilitate the discovery of relevant inter-view relationships. Indeed, the variability in the simulated avatars integrates both subject-specific and population-level (aleatoric) variability. The latter source of variability is captured through the learned variance that is shared across the population. Our goal is to comprehensively assess what the model has learned by introducing subject-level perturbations and investigating their effects on the reconstructed digital avatars. The proposed strategy can be divided into three stages (inference, simulation, and association), which are outlined below and illustrated in Fig. 4.3.

Inference - estimating likelihood distributions

For a given subject $i \in \{1, \dots, L\}$, we propose to use $p_\theta(s^i|z^i)$, the likelihood distribution of observations conditioned on the latent variable learned from the data, to sample DA artificial values $\hat{s}$. In $p_\theta(s^i|z^i)$, θ represents the view l decoder weights, and $z^i = (z^i_{\text{joint}}, z^i_l)$ is the latent representation of s^i, consisting of the view l specific and joint latent representations, respectively (see Section 4.2.2). This likelihood model will tend to sample likely values in the sense of the model. Therefore, provided our model is properly trained, such perturbed values will at best reflect the training data. The estimation of $p_\theta(s^i|z^i)$ is obtained by drawing $D = 1000$ realisations of $z^i \sim q_\phi(z^i|s^i, m^i)$, where ϕ are the weights of the mVAE encoders (see Section 4.2.2). Passing these latent representations to the view l decoder provides a good estimate of $p_\theta(s^i|z^i)$ [197] (by averaging the D decoded reconstructions). Note that the same strategy can be applied to categorical data by approximating the parameters of a Bernouilli distribution instead of a Gaussian.

Simulation - sampling realistic values

Given a subject $i \in \{1, \dots, L\}$, T samples are drawn from $p_\theta(s^i|z^i)$, resulting in T perturbations $\hat{s}^i \in \mathbb{R}^T$. Repeating this sampling for all subjects i, the generated perturbed set forms our DA perturbed view l observations $\hat{s} = [\hat{s}^1, \dots, \hat{s}^L]^\intercal \in \mathbb{R}^{L \times T}$. Importantly, when we generalise to multiple feature in view l, only one feature is perturbed at a time. Finally, the perturbed measures $\hat{m}$ are reconstructed using a forward pass by considering $\hat{s}$ and the corresponding other view's features m as input to the model. This results in a set of T perturbed measures representing our DAs $\hat{m} \in \mathbb{R}^{L \times T \times p_m}$. As a compromise between computational cost and accuracy, we set $T = 200$.

Association - computing inter-view associations

We search for associations between a feature from the view l and each feature from the other view. We refer to the generated DA values for view l as $\hat{s} \in \mathbb{R}^{L \times T}$, and generated DA features for the other view as $\hat{m} \in \mathbb{R}^{L \times T \times p_m}$. Associations are obtained using hierarchical regression models [44]. Specifically, for each feature $j \in \{1, \dots, p_m\}$, a linear regression of the form $\hat{s}^i = c^i_j + b^i_j \hat{m}^i_j + \epsilon^i_j$ is fitted for each subject i. The resulting slopes are averaged over all subjects as $\beta_j = \frac{1}{L} \sum_{i=1}^{L} b^i_j$. Performing the same analysis for all available features of the view l results in the association vector

$\beta \in \mathbb{R}^p$, where $p = p_m * p_s$. This vector encompasses all potential associations between the p_s feature of the view l and the p_m features of the other view.

4.4 . Stability of DAA

We observed that the associations derived from the DAA were not stable when interpreting different models, even when the model was trained on the same training set, and DAs derived on the same left-out set. This instability is likely due to epistemic variability of the mVAE, and has mainly been studied in supervised prediction settings [127, 211]. In feature selection using deep learning, studies have shown that ensembling can enhance stability and mitigate epistemic variability [153]. Starting from these ideas we derive an ensembling procedure adapted to our context and we propose a regularised version of the DAA which is able to overcome this epistemic variability the mVAE.
On the other hand, this setting does not account for aleatoric variability due to the characteristics of the training and left-out samples, and we noticed that it has an effect as well on the derived associations from the DAA. To account for this variability and ensure that derived associations are not specific to a subset of the population, we propose to adapt a stability selection [250] procedure in the following, which only retains associations that are repeatedly selected across different train / left-out splits.

In the following we will detail these two nested mechanisms we use to reach a stable interpretation of what the mVAE network learnt. First the DAA is regularised through an ensembling procedure, then the finally retained associations are selected using the Meinshausen's procedure.

4.4.1 . Regularised DAA

Considering a given set of left-out subjects, the variability in DAA comes from two sources (see Section 4.3). The DAA procedure itself is non-deterministic, due to its inference step which approximates likelihoods and to the sampling in these distributions during the simultation step. However we observed, by repeating the DAA procedure with different random seeds, that these sources of variability are negligible as compared to the variability induced by the considered trained mVAE.

To overcome this sensibility to the training of the mVAE, we propose the regularised Digital Avatar Analysis (r-DAA), obtained by repeating n_E times: train a mVAE on the training subjects, then perform the DAA on the obtained model using the left-out set. Note that the training set and left-out set remain the same throughout these n_E procedures, only the random weights initialisations and batches during training vary. This procedure is illustrated in Fig. 4.4. By subsequently using ensembling, our objective is to identify stable associations from n_E DAAs association matrix $\boldsymbol{\beta} = [\beta^1, \ldots, \beta^{n_E}]^\mathsf{T} \in \mathbb{R}^{n_E \times p}$. Comparable to classical deep ensembling, our approach involves ensembling candidate associations (which correspond to model predictions in supervised settings [211]) proposed by the n_E DAAs, and regrouped in the $\boldsymbol{\beta}$ matrix. The proposed ensembling procedure requires the definition of an aggregation function f and a decision function g. The aggregation

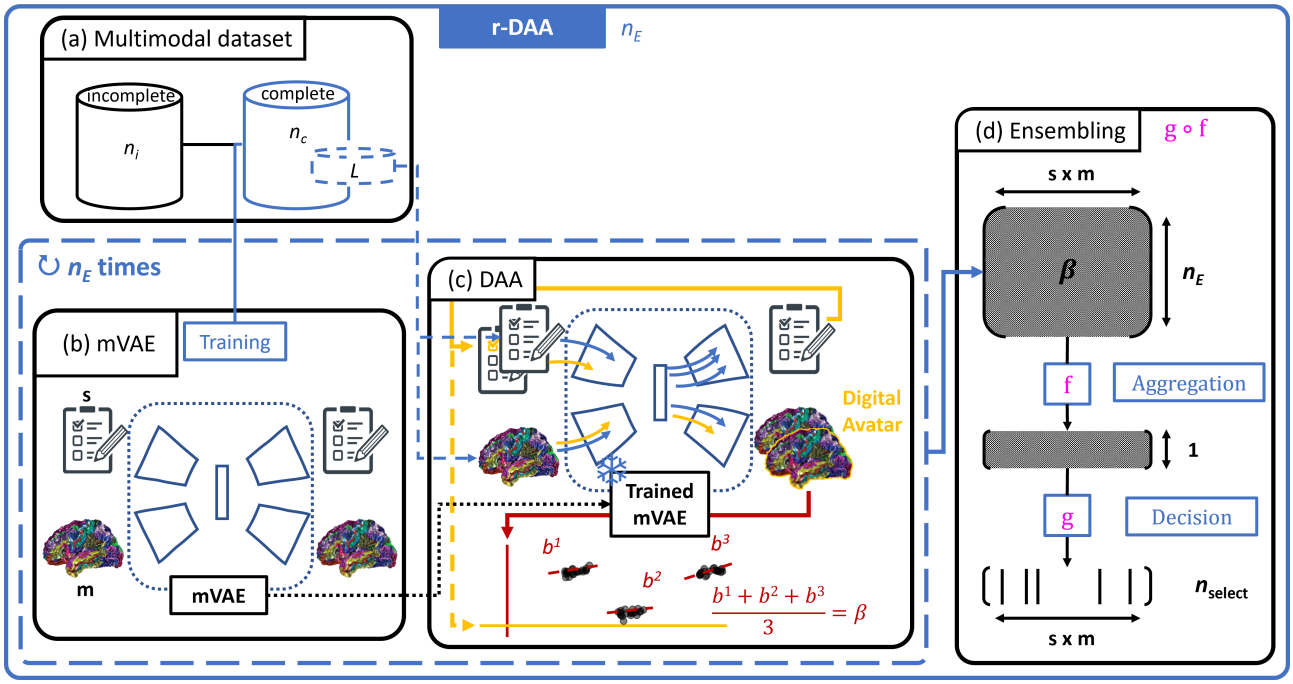


Figure 4.4: Illustration of the regularised Digital Avatar Analysis (r-DAA) in (a) a cohort with two modalities: imaging data (m) and questionnaire scores (s). The (n_c) complete data is split into training (plain blue removing the dotted part) and left-out (dotted blue part) subjects. The r-DAA (blue bounding box) procedure uses a given split of the dataset. In the inner dotted blue bounding box, we repeat the following n_E times: (b) a mVAE is trained using all the (n_i) subjects with one modality and the training subjects with both. (c) The L left-out subjects are used for DAA using this trained model, outputting coefficients β. (d) These n_E DAA outputs are then ensembled with an aggregation and a binary support decision function into n_{select} associations. The r-DAA outputs sparse associations between the questionnaire scores and imaging measures.

function summarises the n_E associations coefficients. The decision function generates a binary decision support from the aggregated coefficients, which is designed to retain meaningful associations. Formally, let $f : \mathbb{R}^{n_E \times p} \to \mathbb{R}^p$ be an aggregation function and $g : \mathbb{R}^p \to \{0, 1\}^p$ be a decision function. The composition $g \circ f$ forms the proposed ensembling, which is outlined below and illustrated in Fig. 4.4(d).

Aggregation - f

The role of the aggregation function f is to assign an importance to each association from the association matrix β. Initially, we opt for a classical mean, although alternative functions such as median or maximum could also be relevant. However, the quality of the estimated n_E models is not accounted for by any of these functions. In fact, some models may converge to local minima of the optimised loss function, resulting in less relevant models and representations. Such behaviour can produce outlier associations that affect the stability of selected associations [153]. Therefore, we employ a weighted average where the weights reflect the quality of the estimated models. The

weights are determined by the model's joint latent space ability to capture a significant proportion of the investigated modality related variability. Practically, we use the Representational Similarity Analysis (RSA) [206], which computes correlations between the joint latent space and the investigated modality features (see Supplementary B for details).

Decision - g

The purpose of the decision function g is to obtain binary decision support from the aggregated coefficients on a feature basis. This support allows us to choose a subset of the most informative associations for each pair of feature (s, m), s being features of the modality investigated against other features m from other modalities, enhancing reproducibility and interpretability. While a simple strategy involves specifying a numeric threshold or defining built-in heuristics to find this threshold, we prefer setting the number of features explicitly rather than using a threshold. This approach provides consistency across score-metric pairs and DAAs. In practice, we choose to select the top n_{select} associations with the greatest amplitudes.

Finally, n_E can be interpreted as a level of DAA regularisation. Increasing n_E should tend to overcome epistemic variability, hence the higher should be the better. But we choose a different framework and heuristic in order to set this n_E regularising strength, the aforementioned stability selection [250]. It features a more flexible selection paradigm which automatically selects the regularising parameter, and it is able to model aleatoric variability as well.

4.4.2 . Stability selection

Stability selection, derived from penalised machine learning [250], is designed to reinforce feature selection through the definition of stability paths. Stability paths represent the probability of selecting each feature when training the same algorithm with some regularisation parameter on different random splits of a dataset. We extend this methodology to our DAA based association study. In our context, the regularisation parameter is the number of models n_E considered in the ensembling step. We propose to repeat the r-DAA introduced in the previous section N times, using different splits of the original dataset, and setting the regularisation parameter n_E in a range of values such as $\{1, \ldots, 20\}$. This approach allows us to model aleatoric uncertainty inherent to population variability. The implementation comprises three aspects outlined below and illustrated in Fig. 4.5. First, we define a valid splitting strategy. Then, we repeat the r-DAA N times by varying the regularisation parameters n_E, enabling us to estimate stability paths. Finally, we define a criterion to assess the stability of the results obtained across the N splits.

Splitting strategy

Our dataset consists of two parts: n_c subjects with complete data and n_i subjects with incomplete data. It's important to note that subjects with missing views can only be used to train the mVAE model. For each split, out of the n subjects available, $n_i + \frac{4}{5}n_c$ are used in the training set. The latter includes the n_i subjects with incomplete data and $\frac{4}{5}n_c$ randomly selected subjects with complete data. The remaining $L = \frac{1}{5}n_c$ subjects with complete data form the left-out set (see Fig. 4.5(a)-(b)). To maintain population statistics, including age, sex, and acquisition site distributions, we employ

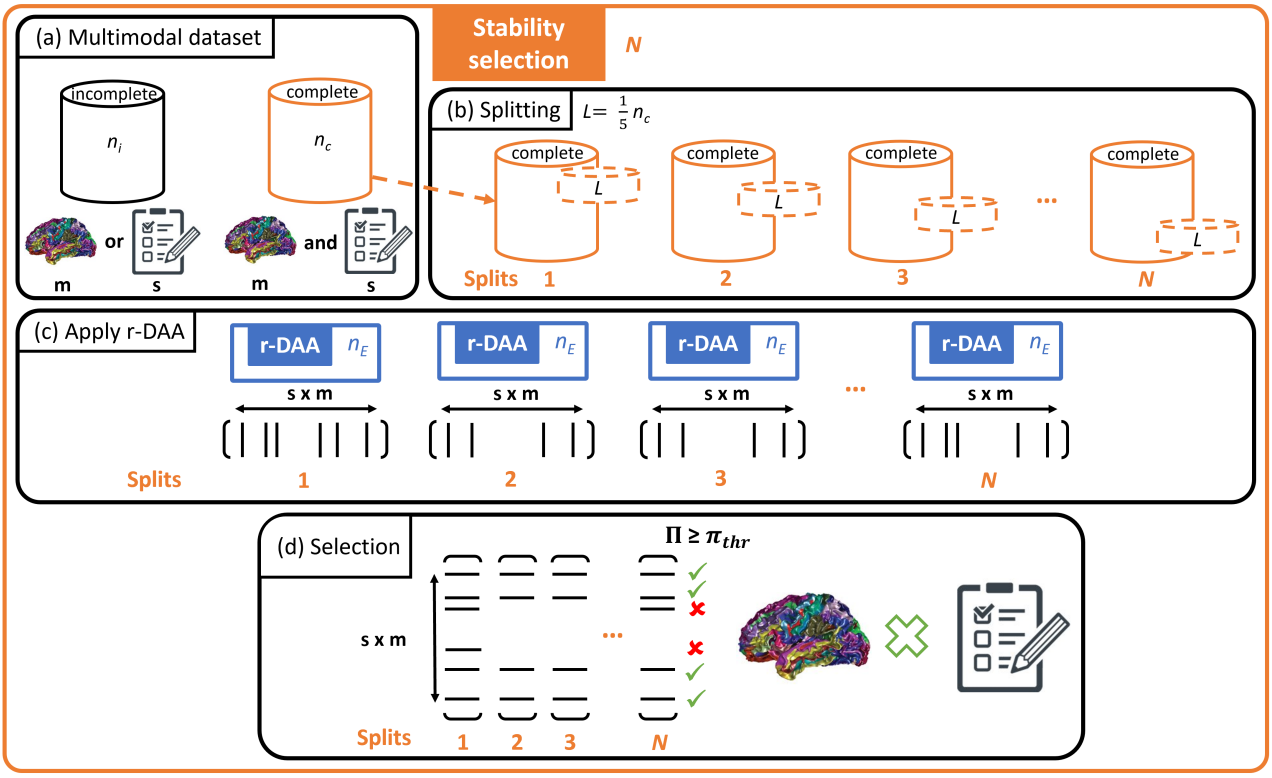


Figure 4.5: Illustration of the stability selection procedure based on DAA in (a) a cohort with two modalities: imaging data (m) and questionnaire scores (s). (b) The complete dataset is split N times into $\frac{4}{5}n_c$training (plain orange removing the dotted part) and $L = \frac{1}{5}n_c$ left-out (dotted orange part) subjects. (c) The r-DAA procedure is applied to each split using all the dataset, outputting sparse associations between the questionnaire scores and imaging measures. (d) Stability paths Π are computed for each value of n_E and associations with stability path $\Pi \geq \pi_{\text{thr}}$ are selected. The stability selection outputs stable associations between questionnaire scores and imaging measures.

shuffled iterative stratification [311]. Although using the entire incomplete dataset in each training iteration may appear as a limitation, it is effectively mitigated by employing different shuffled batches at each training epoch. We define N as the number of splits.

Stability paths

Each r-DAA with regularisation parameter n_E produces a decision support $S^j(n_E) \in \{0,1\}^p$ for a given split $j \in \{1,\ldots,N\}$. The stability paths $\Pi^{n_E} \in [0,1]^p$ are the probability for each association to be selected across the N splits. It is obtained by averaging all the calculated binary decision supports as:

$$\Pi^{n_E} = \frac{1}{N}\sum_{j=1}^{N} S^j(n_E) \tag{4.3}$$

Stability criterion

The stability criterion analyses the stability paths and defines which associations are considered stable. If the probability of an association happens to be greater than a user-defined threshold $\pi_{\text{thr}} \in [0, 1]$ for a specific regularisation parameter n_E, then the corresponding association is selected. Finally, we define the set of stable features as:

$$S^{\text{stable}} = \{k : \max_{n_E \in \{1,\dots,20\}} \Pi_k^{n_E} \geq \pi_{\text{thr}}\} \tag{4.4}$$

More simply, this means that associations with a high probability of selection for any regularisation parameter are retained as stable associations. Conversely, those with a low probability of selection are dropped.

4.5 . Conclusion

In this chapter, we introduced the need for integration methods that can model multi-view data acquired on a single subject, as they are available in population studies. These population studies are widely used to study psychiatric illnesses about which little is known. These cohorts are known to be heterogeneous, not only because it is difficult to control diagnoses precisely during the inclusion, but also because these diseases display comorbidity. Following criteria such as RDoC, these population studies consist of multi-modal acquisitions (e.g. imaging, genotyping and clinical scores) to find "dimensions" underlying the symptoms or psychiatric disorders. Integration methods are required to handle multiple biological markers and model their contribution to multiple dimensions of pathology, that can be shared across psychiatric disorders and possibly contribute to non-pathological variability. We considered a number of properties required by machine learning models to address this problem.

We introduced candidate integration models considering linear approaches, which include CCA, PLS and their extensions. However, their lack of capacity and inability to use observations with missing assessments are very limiting.
Deep PGMs present very appealing properties, as they are able to leverage observations with missing modalities and can model the correlation structure intra (with modality specific networks and private latent spaces) and inter-modality (with joint latent space) efficiently. Additionally, they have been shown to embed disentanglement properties, which can be used to separate between view-specific and shared variability. This can have important implications, as most biases are not contained in all modalities: image acquisition site effect is a very good example of such bias, which will be restrained to imaging modalities. Sex-related effects might be restrained to imaging and genetic modalities. Clinical questionnaires are designed to be unbiased by such factors.

Then we discussed the interpretability opportunities which come with the models introduced previously, and propose an interpretation method using Digital Avatars which interpret what a model learned by slightly modifying features of one modality in input space and interrogating the mVAE to generate the corresponding changes in the other modalities. This interpretation module outputs an association matrix linking each feature from each modality to the features from the initially perturbed modality. It provides a convenient interpretation module to understand the inter-modal

relationships learned by a mVAE model.

Additionally, we identify stability as key to enhance the reproducibility of findings. After discussing some origins of instability in ML, we propose ways to mitigate these sources of variability when interpreting mVAE models with the DAA. Epistemic variability is handled by repeating the training / interpretation procedure of the mVAE on the same train / left-out split of the population, then ensembling their results. This consists in the regularised DAA. In order to ensure that the exhibited associations are not specific to the train / left-out split considered by the r-DAA, we also propose to use a stability selection procedure to retain consistently selected associations by the r-DAA across multiple splits, hence controlling the aleatoric uncertainty.

These methods were introduced in this chapter in a more general context, but were in fact developed for discovering reproducible brain-behaviour associations. We believe that applying such techniques to large multimodal cohorts could enable finding relevant biomarkers for clinically relevant phenotypes. The DAA led to a publication in the conference ISBI, presented in 2023 at Cartagena, Colombia and during a poster session at the french seminar IABM 2023 in Paris. In the next chapter, we apply the DAA module to a transdiagnostic cohort of at-risk children for various psychiatric diseases (including autism, hyperactivity and anxiety), presenting shared or specific symptoms for these pathologies. We enforce reproducible associations discovery by applying the r-DAA and stability selection procedures. We discuss the related experimental settings and obtained results, as well as some hints on how to interpret them, especially from a transdiagnostic perspective.

5 - Transdiagnostic brain-behaviour study

Contents

Today, psychiatry, in both its diagnostic and therapeutic dimensions, is moving from a paradigm based on the study of syndromes to a new one built on an understanding of their underlying neurobiological mechanisms. Achieving this goal is an ongoing process that requires the combination of scientific advances, technological innovations, and a patient-centred approach [368]. Identifying relationships between behaviours and brain measures is a key aspect of this paradigm. However, behaviours are complex and often result from a combination of genetic, environmental, and psychological factors. Furthermore, the phenomenon of comorbidity frequently discussed in psychiatry means that multiple behaviours associated with various mental disorders can co-occur simultaneously, and the same disorder may manifest with different behaviours in different individuals. This complexity challenges the idea that a single behaviour signature can correspond to a sole disorder. As previously introduced, the NIMH RDoC provides recommendations for properly addressing this complexity. In contrast to traditional approaches that aim to find a diagnosis from a specific score or modality (as outlined in the DSMs), the RDoC promotes dimensional and transdiagnostic approaches. These dimensions include genetic, biological, environmental, and lifestyle factors in the research on personalised psychiatry. In recent years, the transdiagnostic literature in psychopathology has developed along these lines [125]. Some recently developed transdiagnostic approaches investigate a general psychopathological factor, the so-called p-factor [53, 55]. This p-factor would underlie mechanisms common to several psychiatric syndromes and is usually represented as a single dimension that would act on a set of symptoms and, ultimately, disorders. The goal of such studies is to search for neural correlates of the p-factor using various biological markers, such as imaging or genetics.

In a first section, we briefly introduce this transdiagnostic view of psychiatry and shed light on their foundations and possible advantages to find biological markers for psychopathology and treatment. In a second section, we introduce a study of the transdiagnostic cohort HBN, composed of at-risk children for developing psychiatric disorders, presenting symptomatic behaviour. This study offers a stratification of autistic patients within HBN, using multiple questionnaire-based symptom scores. We propose a replication study using a mVAE integration model, with which we are able to replicate similar groups in its latent space. However, due to its central consideration of autism to define subgroups, this study can only reveal transdiagnostic markers related to this pathology. Additionally, imaging data is integrated post-hoc. We use a mVAE to integrate both imaging data and clinical scores during training. However, this mVAE constraints both latent spaces to have same dimension. It prevents it to be applied in settings where views have large dimensional discrepancy. In a third section, we propose to use the methods introduced in the previous Chapter to discover stable brain-behaviour associations. We show the role of each module of our methodological developments in stabilising the output associations, and display these associations in a transdiagnostic perspective. Indeed, these associations are derived from a multivariate low dimensional latent space expressing shared variability between cortical measures and clinical scores correlated with multiple psychiatric disorders. Then, we discuss the presented results, from a methodological point of view, their potential implications in revealing transdiagnostic biomarkers and limitations. Finally, we put them in perspective and propose various possible future works.

5.1 . Shared factors across disorders: transdiagnostic psychiatry

Some research works hypothesise that multiple evidences for shared psychopathologic factors have been ignored too long while investigating and treating psychiatric diseases. The RDoC advocates to take into account transdiagnostic factors to search for biological correlates of psychiatric disorders and design treatments. This line of research is being developed and transdiagnostic factors are being studied, both in terms of hidden biological mechanisms and in terms of possible clinical interventions.

5.1.1 . The transdiagnostic hypothesis

Transdiagnostic research originated from the study and therapy development for treating eating disorders [125]. It was later extended to study anxiety and depressive disorders, then to all psychiatric diseases [53]. The transdiagnostic research relies on two main assumptions: 1) many disorders share common symptom expression, which could be due to an underlying shared aetiology. 2) developing disorder-specific treatments without considering their heterogeneity or high comorbidity is limiting and prevents cognitive behavioural adapted treatments.

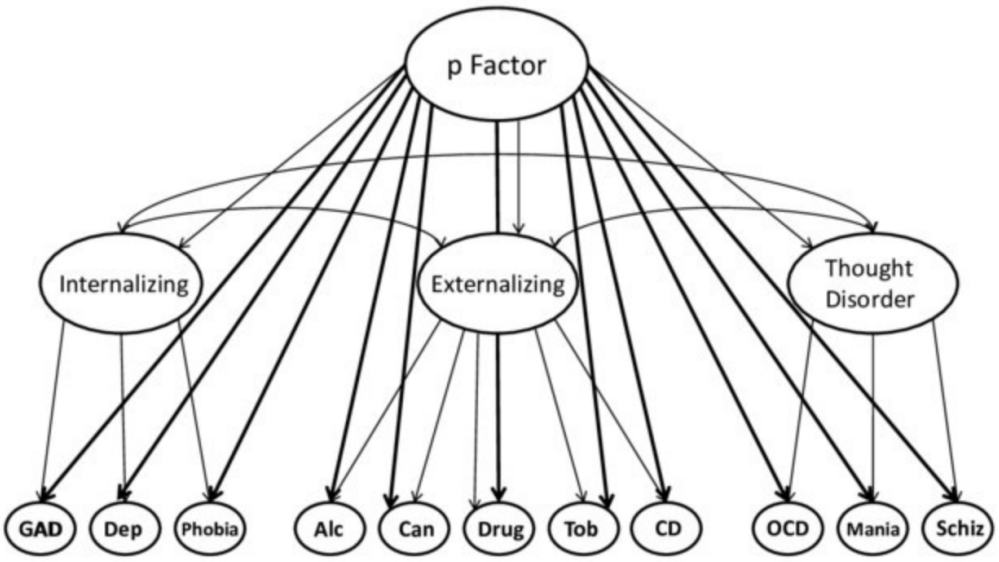


Figure 5.1: Illustration of the hierarchical latent factor model proposed by Caspi et al. [53]. Image taken from [249].

In psychiatry, an emerging concept is the general psychopathological p-factor [53]. It proposes a hierarchical latent factor modelling (illustrated in Fig. 5.1) of psychiatric disorders. The second level of this model would be composed of already identified dimensions underlying most psychiatric disorders:

- the **Internalising** dimension reflects susceptibility to present symptoms related to mood disorder or anxiety, such as major depression, generalised anxiety, panic disorder and social phobia.

- the **Externalising** dimension would underlie liability to substance abusing and addiction behaviours and antisocial attitudes.
- the **Thought Disorder** spectrum assesses sensitivity to psychotic thinking and behaviour such as schyzotopal personality disorders.

The first level of this hierarchical model would be composed of the **General Psychopathological p**-factor, proposed as another latent dimension underlying both psychiatric disorders and these 3 psychopathological processes. It would unify their different yet overlapping theories and explain their correlation. This work aims at overcome DSM diagnoses by proposing a novel theory more suited to take into account comorbidity and shared etiology across disorders. From this pioneering proposition, psychiatric research has started to take into account such factors when designing therapeutic treatments.

5.1.2 . Transdiagnostic applications

We can distinguish two main types among trandiagnostic studies. On the one hand trandiagnostic processes are investigated, proposing different modelling as the one introduced by Caspi et al. [53], and neural correlates are investigated using different imaging techniques. On the other hand, transdiagnostic clinical interventions are being explored.

Transdiagnostic processes study. A question that concentrated some research effort was the statistical latent factor model which should be considered in transdiagnostic settings. It was in fact shown that this modelling was not the most important research question, and that multiple models give reasonably similar results [54]. As such, research should be focused at finding biological correlates of such p-factor rather than statistical modelling, and proposing well-specified validation criteria. So far, most transdiagnostic studies investigating biological processes do not fulfil these criteria [125, 87, 257]: most studies have large discrepancies in experimental settings, considered population, diagnostics and validation. Specifically, only a few studies replicate in independent external cohorts.

Transdiagnostic treatment. Treatments in coherence with this transdiagnostic hypothesis allow to focus on developing a unified manner to treat patients with psychopathologies. Many treatments have been explored thus far and seem to show encouraging results improving over conventional approaches (derived from case / control studies), particularly in treating anxiety and depression [309] and comorbidity [87]. Transdiagnostic interventions include moslty two broad treatment categories: universal interventions promote a one-size-fits-all protocol where all patients receive the same therapeutic treatment. Other approaches propose a more modular treatment, composed of evidence-based functional units (modules) that can be delivered flexibly in a more personalised way.

The HBN cohort [6] is a large multi-center, transdisciplinary clinical study. It includes a variety of assessments, including imaging, and a comprehensive set of psychological and clinical assessments to better understand psychiatric disorders. Inclusion criteria are not diagnosis dependent, but rather encompass an at-risk population with notable behavioural symptoms. Specifically, subjects

were selected based on the presence of behavioural constructs related to ASD, ADHD or other anxiety disorders. Consensus diagnoses are not available for the majority of subjects enrolled in the HBN cohort. As such, the dataset allows for the study of the different manifestations of psychiatric syndromes within the data. In particular, it provides an opportunity to explore methods for new biomarker discovery and dimensional analysis. In this regards, HBN offers a very promising candidate dataset for transdiagnostic studies to identify biomarkers underlying biological processes across disorders.

Now, we will present two studies using HBN which integrate such a dimensional strategy to identify transdiagnostic correlates between brain and behaviour. These works differ in their central focus, with the first study concentrating on the cortical signatures of autistic subgroups, while the second directly investigates transdiagnostic associations as part of a more integrated procedure. A particular attention is paid to assess or improve reproducibility of these studies.

5.2 . Cortical signature of autistic subgroups in HBN

5.2.1 . The dataset

This HBN release we used contains approximately 1800 subjects. Considering participants with overlapping available behavioural assessments and with FSIQ > 70, 1093 participants remain. More precisely, 7 behavioural scores were used: a quantitative measure of clinical autistic traits was defined as the parent Social Responsiveness Scale (SRS), hyperactivity levels were determined using the hyperactivity subscale within the Strengths and Difficulties Questionnaire (SDQ-ha), the anxiety was measured using the total score from the Screen for Child Anxiety Related Disorders Parent-Report (SCARED), irritability was defined using the total score of the Affective Reactivity Index Parent-Report (ARI), and finally, levels of depression, aggression, and attention problems were determined using subscales of the same names within the Child Behavior Checklist (CBCL-wd, CBCL-ab, and CBCL-ap, respectively).

T1-weighted MRI available for these 1093 subjects were used, resulting in 527 images after preprocessing and QC, acquired at 3 different sites mostly at 3T and some of them at 1.5T. They were preprocessed using FreeSurfer to extract cortical thickness, surface area and local gyrification index and represent them onto the fsaverge template of order 7 (~ 160K vertices per hemisphere). QC include manual inspection and controlling for the reconstructed surface (Euler number < -217).

From this dataset, we introduce a previous analysis where an expert derived 3 autistic subgroups in a dimensional approach, leaning on an unsupervised clustering algorithm applied to these 7 clinical scores. These clusters are then qualified and further described as 3 types of autism. The groups are compared to each other, and with other participants considered as controls, in terms of measures of their cortical surface, in an attempt to find biological signatures specific to each of these subgroups. In a second time, we present a study that uses very similar data from HBN and replicates some of the previous findings, using a completely different technique. We used a particular mVAE, the sMCVAE, to integrate both questionnaires and imaging data at training time. By inspecting this sMCVAE latent

space, we remark that the participants are grouped with the same stratification proposed by the previous study [253]. We further try to interpret the network using its generative abilities.

5.2.2 . Expert stratification assisted by unsupervised clustering

This study [253] focuses on finding a stratification among the autistic population within HBN. An unsupervised algorithm is used to determine clusters of subjects in the HBN cohort. Then a stratification of subjects with autistic profiles is derived by selecting 3 subgroups that are further characterised by their symptoms and cortical differences.

The stratification of the 1093 participants was done using a K-Means clustering on the 7 z-scored behavioural scores. The optimal number of clusters was adjusted using the Bayesian Information Criterion (BIC).
The cortical features were harmonised for site effect using the ComBat correction [121], then smoothed using gaussian kernels. Statistical analyses of the cortical measures were done using a linear mixed model to model group effect for each of the autistic subgroups, adjusting for sex, age and FSIQ.

The BIC criterion identified 9 as being the optimal number of clusters in this population. Out of them, 3 presented relatively high mean SRS. As part of their expert analysis of these 9 groups, the authors chose to retain 3 clusters whose subjects show an average SRS score above 80, and to study the clusters that partitioned these subjects. For each of these 3 autistic subgroups, the average values of the 6 other behavioural scores (other than the SRS) are represented in the radar plot (illustrated in Fig. 5.2):

- The *Emot* (emotional dysregulation) subgroup showed high level of reactivity, aggression and ADHD-like symptoms in $n = 107$ patient.
- A cluster with $n = 82$ subjects maintained normal level in a behavioural symptoms except for attention problems and hyperactivity, thus forming the *Attn* (attention problems) subgroup.
- The last *AnxDep* subgroup with $n = 61$ patients showed high level of anxiety and depression, as well as attention deficits.

The remaining 6 subgroups, with relatively low levels of social impairment assessed by their average SRS, were considered as controls.

Regarding statistical analyses, no significant differences were found when comparing ASD subjects (regrouping the 3 previously found subgroups) to controls. However, some differences where found, when comparing each of these subgroups against controls, which are summarised in the Figure 5.3. These findings tend to highlight that the heterogeneity of the autistic population does not allow to discover biomarkers at the population level but requires finer defined subgroups.

This work, although displaying interesting autistic subgroups, has some limitations. First, it requires a human expert to annotate and choose the 3 subgroups. Then, other criteria could be used

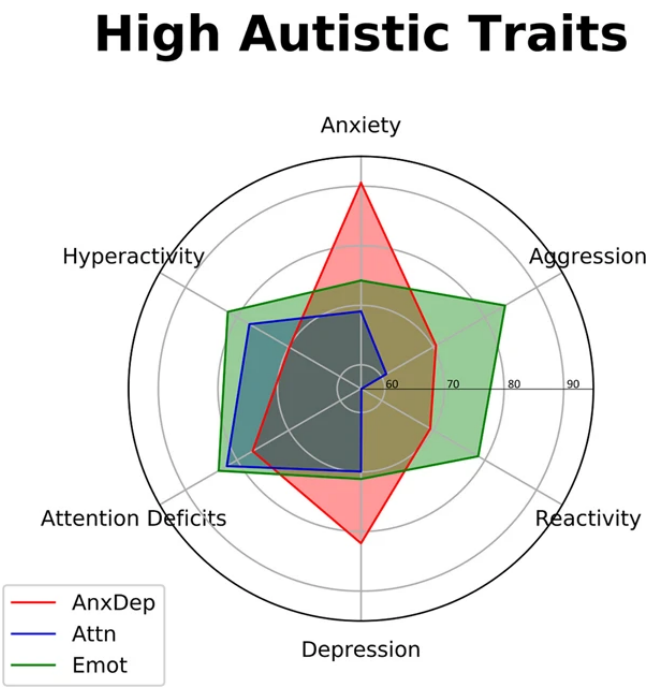


Figure 5.2: Autism subgroups characteristics in HBN found by [253].

to select the total number of cluster, and investigating the stability of these 3 autistic subgroups with different number of clusters in the K-Means could be interesting. And finally, cortical imaging feature were only investigated after having identified the subgroups, when they could have been integrated in the unsupervised algorithm to derive the stratification. This study could benefit from a more integrated perspective. We propose some improvement over this methodological aspect in the study introduced in the next section.

5.2.3 . A replication with DL integration method: sMCVAE

In our study, we use the same dataset and a similar release, but integrate both imaging data and behavioural symptoms at the learning phase using a mVAE called sMCVAE. We use a subset of 619 participants from HBN that have both the 7 symptom scores and imaging data. As in 5.2.2 we selected subjects with FSIQ > 70 and filtered out imaging data preprocessed using FreeSurfer (Euler < -217). We consider the labels defined in the previous study *Emot*, *Attn*, *AnxDep* and controls for all subjects. However, it is worth noting that these labels are solely used for visualisation purpose and are never provided to the learning algorithm.

It must be noted that, instead of considering cortical measures over vertices of the fsaverage template (160K vertices per hemisphere), we use their average value across the 148 ROIs defined by Destrieux's atlas.

Method. The sparse Multi-Channel VAE [14] (see Section 4.2.2) is a particular multi-view Variational Auto-Encoder with modality-specific latent spaces. It does not model explicitly the shared information

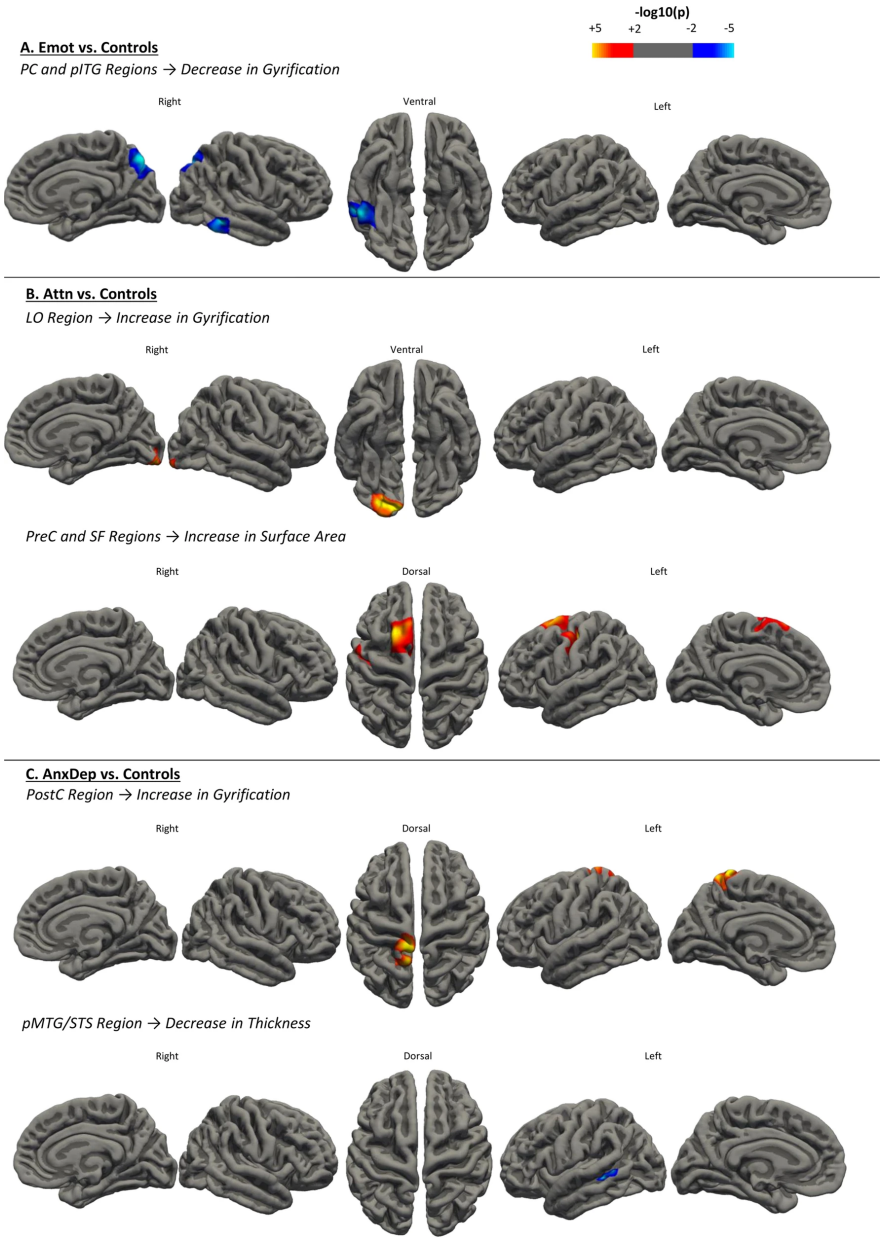


Figure 5.3: Cortical measure differences between ASD subgroups and controls found in [253].

as a PoE-VAE would, but it is trained to reconstruct every view from each latent space. Thus, the latent spaces are constrained to have the same dimension. Moreover, a sparsity penalty can be jointly applied to the different latent spaces to restrain them in using the smallest portion of the space as possible. This sparsity penalisation stems from variational dropout [259], which hypothesises a different form of the latent posterior distribution and introduces and extra parameter $\alpha \in \mathbb{R}^k$, k being the latent space dimension here, which was shown to control the sparsity level of the latent space. This parameter is learned during the training, intervening in the part of the loss involving the

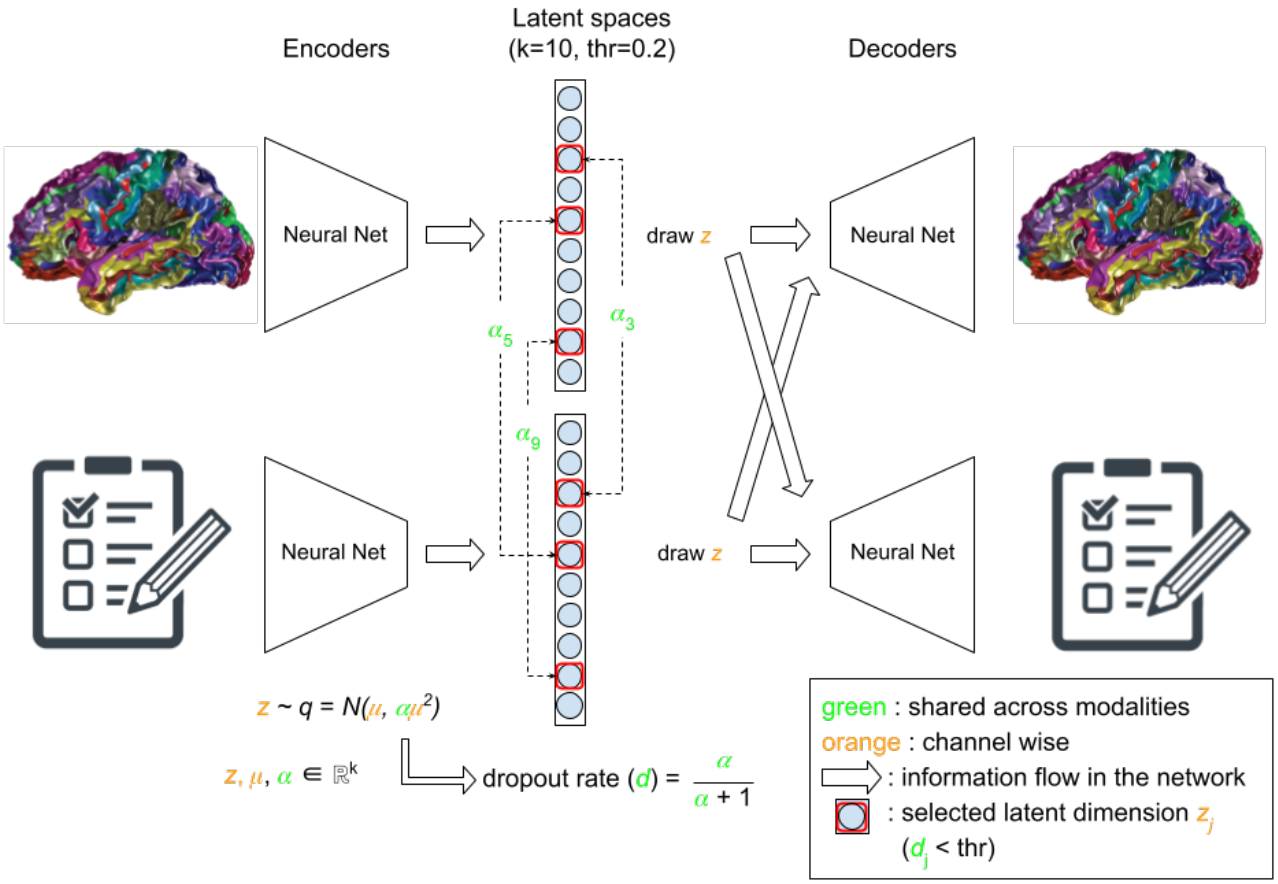


Figure 5.4: Illustration of the sMCVAE used in our reproduction study. Each modality has its own latent space, but both representations are fed the decoders at training time. The approximated posteriors, which have the form $q = \mathcal{N}(\mu, \alpha\mu^2)$, create a dropout mechanism d, function of α, that selects the latent dimension across modalities. The used threshold thr for selecting dimensions with respect to their dropout rate and latent space dimension k are displayed at the top centre of the schema.

Kullback-Leibler divergence. Setting a threshold on the associated dropout rate allows to discard useless dimensions at the end of the training. This model is illustrated in Figure 5.4. We use linear encoders and decoders, and $\frac{4}{5}$ of the available subjects to train the network, while the others are used to assess the generalisability of the trained model. The initial latent space dimension is $k = 10$. This dimension was selected to be smaller than all input dimensions, while knowing that the sparsity parameter would automatically reduce it.

To visualise the differences in cortical measures across subgroups, we consider the centroids of each previously defined subgroups within test participants (not used to train the network) in the latent space of the clinical scores. These centroids are then reconstructed using the imaging decoder and compared to the average Ground Truth (GT) features of subjects from each group. We also compared these reconstructed image features from clinical representations centroids across groups to findings

from the previous study introduced in 5.2.2.

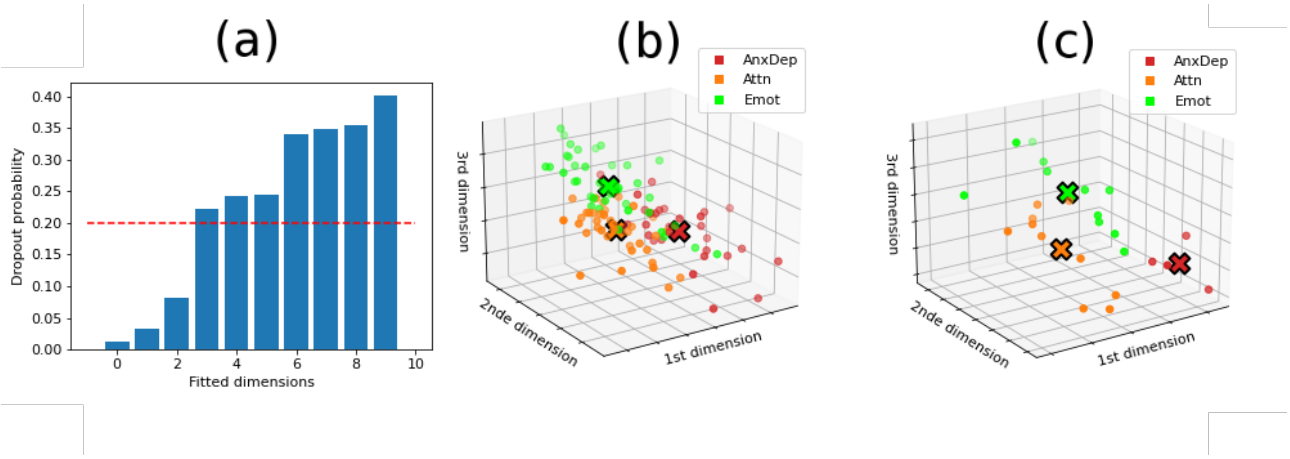
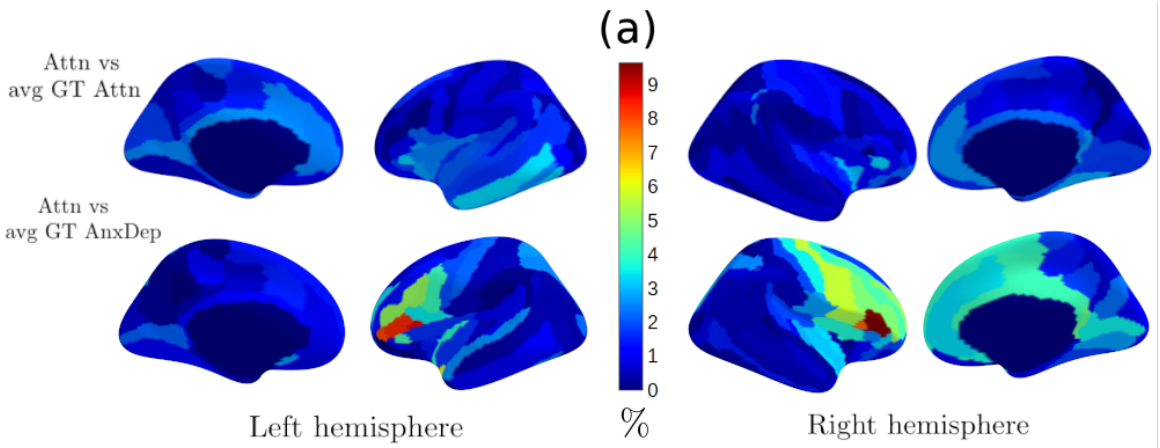


Figure 5.5: Highlighted trained sMCVAE latent space properties. (a) Sorted dropout probabilities of the latent space dimensions, and the threshold used as a red dotted line at $d = 0.2$. The clinical latent space is displayed with the three retained latent dimensions. Participants labelled as autistic in [253] are displayed, with their subgroup labels. Each subgroup centroid is represented by a cross. In plot (b) are represented training subjects and test subjects are represented in plot (c).

Results. During training, the dropout rate optimised by the sMCVAE converged for the different dimensions: among the 10 initially deployed latent dimensions, 3 of them were identified as relevant by this sparsity constraint, considering the advocated dropout rate threshold of 0.2 by the authors. This allowed us to visualise directly the subjects representations in the latent space, without relying on an additional dimensionality reduction such as PCA or t-SNE. Figure 5.5 displays information about the learned latent space. In Fig. 5.5(a), we can see the sorted dropout rates stemming from the learned α configuring the posterior latent distributions. Their relationship is illustrated in Fig.5.4. The displayed threshold was used to discard latent dimensions with dropout rate above it.

Figure 5.5(b) and (c) display the thus learned 3 dimensional clinical latent space, corresponding to the output from the clinical encoder. Training subjects representations are plotted in this latent space in Figure 5.5(b). Note that only non-control subjects in the previous work, coloured along with their subgroup labels, are displayed here for clarity. Subgroups appear to be smoothly regrouped by labels, although there is some overlap. In Figure 5.5(c) are similarly displayed the clinical representation from the test subjects. The groups appear even more broadly separated, showing that the network properly generalises to unseen data.

Additionally, the reconstructions from centroid clinical representations show relatively low errors, as shown in Fig. 5.6(a), when compared to their average GT. When comparing to average GT of other subgroups, errors increase in some areas. These areas can in fact highlight potential ROIs for differencing the subgroups. Note that statistics regarding differences in cortical measures between ASD



(b)

Type of feature	subgroups : ROI	Controls	Attn	AnxDep	Emot
Local Gyrification Indice	rh:g:precuneus	3.185	3.188	3.118	3.181
	rh:g:parietal:sup	3.251	3.244	3.176	3.250
	rh:g:temporal:inf	2.852	2.844	2.804	2.852
	rh:g:temporal:middle	3.629	3.630	3.548	3.625
	rh:s:parieto:occipital	3.555	**3.569**	3.480	3.548
Surface Area	lh:s:central	2255.205	**2282.031**	2225.467	2247.934
	lh:s:precentral:inf:part	1224.380	**1241.927**	1191.967	1218.565
	lh:s:precentral:sup:part	1022.335	**1032.413**	994.600	1018.587
	lh:g:precentral	1771.069	**1793.433**	1723.102	1763.518
	lh:g:front:sup	5239.351	**5337.183**	5159.042	5210.950
	lh:s:front:sup	2249.695	**2288.085**	2196.221	2238.227
Thickness	lh:g:temporal:middle	3.310	3.314	**3.280**	3.312
	lh:s:temporal:sup	2.750	2.749	**2.715**	2.752

Figure 5.6: Reconstruction properties from test subgroup centroid clinical representations. (a) Reconstruction error when compared to the average GT gyrification indices for different subgroup. (b) Reconstructed values for ROIs identified in [253].

subgroups were not conducted in the previous study. However, we can investigate the reconstruction of cortical features from clinical latent representations of each subgroup centroid and compare them to reconstructed cortical feature of the control group. These results are displayed in Figure 5.6(b) and reproduced findings were highlighted in red: the *Attn* subgroup displays increased average surface area in superior frontal and precentral regions of the left hemisphere, and *AnxDep* subgroup presents decreased average thickness in the middle temporal gyrus and superior temporal sulcus of the left hemisphere. These findings were however not statistically validated.

5.2.4 . Discussion

In this section, we discuss the two presented analyses in the two previous sections, their advantages and limitations, and their pertinence for our research question, which is finding transdiagnostic brain-behaviour associations.

Reproducibility of the results. The results obtained in the study presented in Section 5.2.3 reproduce the outcomes of the paper from [253]. Indeed, we find a similar stratification of the clinical

latent space in our trained mVAE, without ever using these labels to train the network, although we can note various differences in the data and modelling process we used:

- We used a slightly different subset of HBN with a few more images (619 vs 527 subjects).
- We used measures averaged over Destrieux's ROIs instead of considering the spherical template fsaverage.
- We did not use a smoothing kernel over the cortical surface data, because averaging over ROIs did the work for us.
- We did not control for site effects.
- We trained an integration model using both clinical and imaging data.

This is encouraging as reproducing results in such a different setting is often hard. However, data were preprocessed by the same people and the same software, which surely helps a lot.

Regarding the imaging results, we can not really pronounce about replicability. First, imaging features were integrated in the learning algorithm, as opposed to post-hoc analyses in the previous study. Moreover, the proposed analyses in our replication study were very shallow and preliminary. Validating them would require further experiments and statistical validations. It is also difficult to compare them, as the considered imaging features are different, and the smoothing effect coming from averaging over ROIs could hide small but significant effects.

Focus on autism. The labels considered in these studies rely mostly on the SRS clinical scale and participants with lower SRS are directly labelled as controls. This assumption is certainly better to study ASD in a more dimensional approach than binary DSM diagnostic, but is quite limiting in the transdiagnostic case.

Moreover, the stratification proposed in the study presented in Section 5.2.2 is completely driven by behavioural scores. A more integrated approach to stratification could be used, to take into account both symptom scores and biological variations reflected in the images. For instance, the latent clinical latent space of the trained sMCVAE is entirely determined by clinical scores, since it is uniquely generated using the clinical modality. That being said, since these representations are passed to the imaging modality decoder during training, the clinical latent space is organised in such a way that imaging features can be reconstructed from the clinical latent space as well (such as in Figure 5.6). It is therefore a good candidate space for an integrated stratification.

About the sMCVAE. From a methodological perspective, we can now discuss pros and cons of using the sMCVAE.
The sparsity constraint is very interesting and works nicely. It constraints the latent space to remain sparse and regularises the network, which has the potential of limiting the number of solutions when considering more complex and non-linear NN architectures such as MLPs. However, this sparsity acts as a proportion regularising constraint, not so much as a number of ideal dimensions seeker. Indeed if you increase the network latent dimension (say 100 instead of 10), it will select the same proportion of this novel latent space (here 30 instead of 3) rather than a fixed

number of dimensions. Additionally, this sparsity mechanism dictates that the network must have the same latent space dimension for all modalities. In the case of MCVAE, this constraint already arises from the fact that decoders are given all latent representations. But this can be problematic in settings where there is a large dimensional discrepancy between modalities. The final latent space dimension should be less than input dimension, otherwise the network can encode any information, including hard-coding the data, which is not desirable. Having flexibility in setting different latent space dimension in multimodal integration using DL is of major importance.

Interestingly, not residualising for site effect did not impact (or only slightly) the clinical latent space. This highlights this convenient property arising from modelling modality-specific latent spaces, i.e. that are not computed using other modalities. However, this setting will never allow for a proper disentanglement of shared and modality-specific factor, because of the way it is trained. This separation will only happen if an explicit joint latent space is modelled separately.

In this study, we only used a linear encoder and decoder, which limits the number of solutions and guaranties a reproducible network throughout different trainings (less prone to epistemic variability). However, it does not explore NNs capacity to their full extent, and is not very realistic to integrate larger dimensional modalities with complex structure such as cortical surface data on the fsaverage.

Overall, this study is a great example of reproducible stratification, but its replication in other cohorts is still unclear and the discrepancies in clinical questionnaires and recruitment criteria across cohorts are very hard to come by when considering such a stratification strategy. In the following, we focus on finding brain-behaviour associations, that we deem to have a better potential for replication in other cohorts. These associations are mostly driven by symptoms (using symptom score), and should thus not be limited by recruitment criteria.

5.3 . Towards reproducible transdiagnostic brain-behaviour associations

The previous study focuses mostly on ASD. In the following, we want to discover reproducible transdiagnostic associations between measures of the cortical surface and symptomatic behaviours implicated in multiple psychiatric disorders. Here, we showcase the capacity of the tools previously introduced in Chapter 4 to uncover such stable associations. We integrate questionnaire scores and cortical measurements over ROIs using a mVAE, then use the DAA framework to interpret its joint representations and output associations. The r-DAA procedure is used to extract associations that are not due to epistemic variability, wrapped in a stability selection framework which ensures that retained associations are generalisable within our population, and therefore more prone to being reproducible.

5.3.1 . The dataset and parameters of the methods

Multimodal dataset. This study uses the HBN cohort as well, but a more recent release containing more participants. In the previous study introduced in Section 5.2.2, our group identified seven

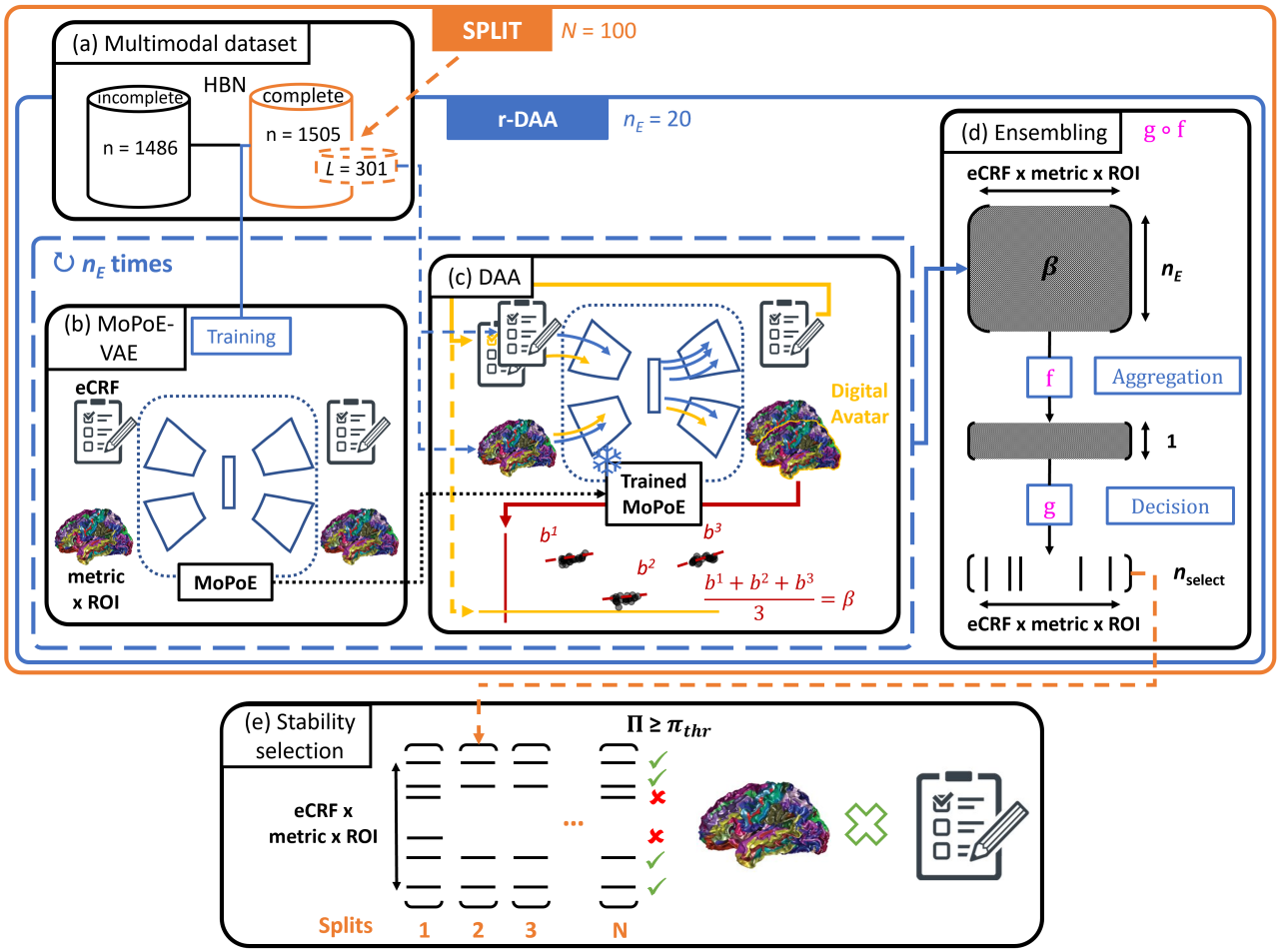


Figure 5.7: Illustration of the interpretation framework based on DAA. In (a) the HBN cohort with two modalities: imaging data (metric $\times$ ROI) and clinical questionnaires (eCRF). For each split (orange bounding box) of the dataset, complete data are split into training (plain orange removing the dotted part) and left-out (dotted orange part) subjects. Each r-DAA (blue bounding box) procedure uses one split of the dataset. In the inner dotted blue bounding box, we repeat the following n_E times: (b) a MoPoE-VAE is trained using all the subjects with one modality and the training subjects with both. (c) Left-out subjects are used for DAA using this trained model, outputting coefficients β. (d) These n_E DAA outputs are then ensembled with an aggregation f and a binary support decision function g into n_{select} associations. The r-DAA procedure is repeated for each of the $N = 100$ different splits of the complete dataset. (e) A stability selection procedure is used to retain associations with stability path $\Pi \geq \pi_{\text{thr}}$ high enough. The whole framework outputs stable associations between clinical questionnaires and imaging measurements on ROIs.

behavioural assessments that capture the most salient dimensions in ASD patients [253]. Specifically, a quantitative measure of clinical autistic traits was defined as the parent Social Responsiveness Scale (SRS), hyperactivity levels were determined using the hyperactivity subscale within the Strengths and Difficulties Questionnaire (SDQ-ha), the anxiety was measured using the total score from the

Screen for Child Anxiety Related Disorders Parent-Report (SCARED), irritability was defined using the total score of the Affective Reactivity Index Parent-Report (ARI), and finally, levels of depression, aggression, and attention problems were determined using subscales of the same names within the Child Behavior Checklist (CBCL-wd, CBCL-ab, and CBCL-ap, respectively). These scores reflect various symptoms implicated in multiple psychiatric disorders, and are thus good candidate to build transdiagnostic factors. In total, 2454 subjects with all these scores are retained.

For this release of HBN, the MRIs were acquired at four sites. A mobile 1.5T Siemens Avanto on Staten Island, a 3T Siemens Tim Trio at the Rutgers University Brain Imaging Center, and 3T Siemens Prisma at the CitiGroup Cornell Brain Imaging Center and at the CUNY Advanced Science Research Center. Collected T1-weighted images were preprocessed using FreeSurfer. The Euler number is used as a quality metric summarising the topological complexity of the reconstructed cortical surfaces [301]. Specifically, a single Euler number exclusion threshold of -217 is applied to yield 2042 selected subjects. Finally, cortical measures based on three metrics - cortical thickness, curvature, and area - are averaged in the 148 ROIs defined by Destrieux's parcellation [96].

The proposed brain-behaviour integrative analysis considers these two blocks of data. First, the electronic Clinical Record Form (eCRF) view consists of the $p_{eCRF} = 7$ behavioural scores. Second, the ROI view is composed of the $p_{ROI} = 444$ cortical features from the 3 considered metrics across the 148 ROIs. In total, our dataset comprises 2991 subjects. Among them, 1505 have both complete views (referred to as complete dataset in Fig. 5.7(a) and 1486 have only one of the two views available (referred to as incomplete dataset in Fig. 5.7(a). Missing views remain a common problem in data integration. The factors contributing to missing data are usually not known in advance. Most traditional data mining and machine learning approaches operate on complete data and fail when data is missing. As fallback strategies, some models rely only on samples with all views available or on an auxiliary inference step that generates missing views. Our goal here is to use a model able to accommodate missing views, in order to use a maximum number of available samples.

Parameters of the methods. We describe precisely in the following the implementation of the notions described in the previous chapter and illustrated in Fig.5.7.

The selected model. We identify the MoPoE-VAE (see Section 4.2.2) as a good model to properly build modality-specific and shared latent spaces, as well as for its straightforward way of handling missing modalities. As a reminder, the MoPoE-VAE models the joint posterior as the mixture of products of (subsets of available) experts. The encoders with parameters ϕ are each defined as MLPs, each with one hidden layer of 256 units and a ReLU activation. The decoders with parameters θ are linear (a single fully connected layer). Note that every decoder has a learnable variance parameter for each reconstructed feature that does not depend on the input observation. This allows the decoders to learn a population-level (aleatoric) variability. To handle inputs of different sizes, the dimensions of the view-specific latent spaces were set individually. Based on a previous study [11], we chose $d_{eCRF} = 1$ and $d_{ROI} = 20$. Defining the dimension of the shared latent space jointly with the dimensions $(J_1, \ldots, J_K)$ of the input views seems to be a reasonable criterion. A

latent space should be smaller in dimension than the input. We use the following rule:

$$d < \min_{k \in \{1,\dots,K\}} \{J_k - d_k\}$$

To spread the variability of the observations across the view-specific and shared latent representations, we choose $d = 3$. All data are z-scored before being fed to the network. The model is trained with an Adam [195] optimisation, an initial learning rate of 2×10^{-3} and a batch size of 256.

Regularised DAA procedure. Considering a training / left-out split of HBN, we apply our r-DAA procedure (see Section 4.4.1), with $n_E \in \{1, \dots, 20\}$: we train n_E models on the training subjects, using subjects with complete data (eCRF and ROI) and subjects with incomplete data (eCRF or ROI), different random initialisations and training batches. From of these n_E trained models are derived associations using DAA on the left-out subjects with complete data, that are concatenated in the association matrix $\boldsymbol{\beta}$. We consider two ensembling functions f in order to compare their results. The first one is the uniformly weighted average function, and the second is an average weighted by a quality measure of our trained MoPoE-VAEs. A model is rated using the Kendall τ correlation between its joint representation of left-out subjects and each eCRF score individually, then averaged across scores (see Appendices B and C.2 for more details). Final associations retained are selected using the latter ensembling function. The decision function g selects the $n_{\text{select}} = 12$ associations of greatest aggregated amplitude $f(\boldsymbol{\beta})$. This procedure outputs a sparse vector $g \circ f(\boldsymbol{\beta}) \in \{0, 1\}^p$ with $p = p_{eCRF} * p_{ROI}$.

Stability selection. We implement the introduced stability selection procedure in Section 4.4.2. Our HBN dataset consists of 1505 subjects with complete data and 1486 subjects with incomplete data (see Section 5.3.1). Recall that subjects with missing views can only be used to train the model. For each split, out of the 2991 subjects available, 2690 are used in the training set. The latter includes the 1486 subjects with incomplete data and 1204 randomly selected subjects with complete data. The remaining $L = 301$ subjects with complete data form the left-out set (see Fig. 5.7). To maintain population statistics, including age, sex, and acquisition site distributions, we employ shuffled iterative stratification [311]. We opt for using $N = 100$ splits.

5.3.2 . Models inspection

To evaluate the information encoded in the various latent spaces of the trained models, we employ the RSA [206] tool introduced in Appendix B. The RSA results in Table 5.1 depict the relationships between different latent spaces (eCRF, Joint, and ROI) and each eCRF score, alongside other relevant covariates like age, sex, and image acquisition site. In short, we computed each of these Kendall τ correlations for every $N = 100$ splits and corresponding $n_E = 20$ models (using left-out subjects). The reported correlations $\bar{\tau}$ are averaged across these $N * n_E$ values and statistically significant correlations are displayed in bold. These reported correlations illustrate how the different latent spaces co-vary with the compared features. See Appendix B for further details.

Notably, each eCRF score strongly correlates with the eCRF-specific latent space, as highlighted in pink in Table 5.1. This is noteworthy since the models successfully learned a one-dimensional space capturing a significant amount of variability associated with all the eCRF view scores. This outcome is

	eCRF	Joint	ROI
Score	$\bar{\tau}$ (↑)	$\bar{\tau}$ (↑)	$\bar{\tau}$ (↑)
SRS	**0.302**	**0.018**	-0.003
SCARED	**0.101**	**0.032**	-0.004
ARI	**0.256**	**0.05**	**0.017**
SDQ ha	**0.326**	**0.058**	0.005
CBCL ab	**0.406**	**0.03**	0.008
CBCL ap	**0.443**	**0.026**	-0.002
CBCL wd	**0.152**	**0.06**	0.008
Site	0.004	0.011	**0.156**
Age	-0.005	**0.05**	**0.10**
Sex	0.014	0.003	**0.067**

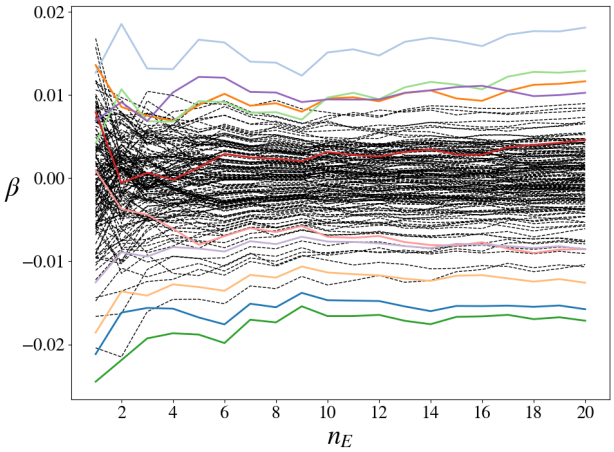
Table 5.1: Representational similarity analysis results between the different latent spaces (eCRF, Joint, and ROI) and the clinical eCRF scores, and some covariates of interest: the imaging acquisition site (Site), the patient age (Age) and sex (Sex). $\bar{\tau}$ is the corresponding average Kendall τ across models and splits. We computed the p-value associated with this statistic. Values in bold indicate their significativity, i.e. median corrected p-value < 0.01, see Supplementary B for details.

expected, considering the correlations between the eCRF scores (see Appendix Fig. B.1). Additionally, it underscores that this latent space is not informing about age, sex, or image acquisition site. In contrast, the latent space specific to the ROI view does not significantly correlate with the eCRF scores with the only exception of ARI, yet with a small correlation. Note that it significantly correlates with age, sex, and image acquisition site, as shown in the cells highlighted in orange. Finally, we highlight in salmon in Table 5.1 what the joint latent space has learned. These representations moderately and significantly correlate with each score but not with acquisition site or sex. Moreover, it appears to correlate with age. It seems that the brain-behaviour relationships, modelled in the MoPoE-VAE joint latent space, relies on information related to age.

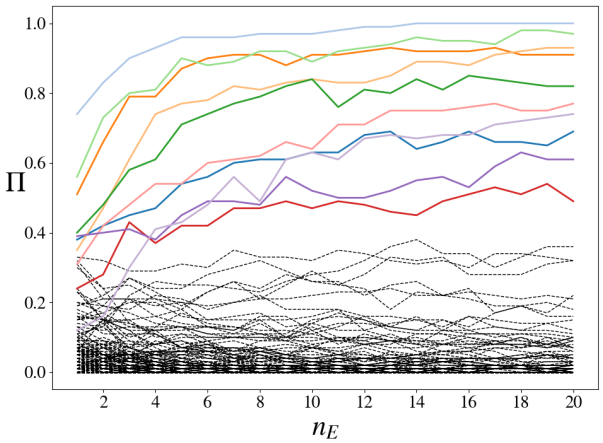
5.3.3 . Stability selection of r-DAA associations

By increasing the number of models and applying a stability selection procedure, we aim to increase the stability of the sets of brain-behaviour associations supported by our multimodal dataset. In the following, we consider the SRS score associations with the thickness metric. Figure 5.8 illustrates the associations trajectories taken by the considered 148 ROIs. This illustration is inspired by the figure of merit proposed in the seminal paper [250]. In Figure 5.8(a), we display trajectories of aggregated β values (i.e., effect sizes), using a mean function for f, against the number of model n_E. Some trajectories become more prominent (display larger amplitude of average effect size) than others as the number of models increases. However, they are hardly distinguishable from other

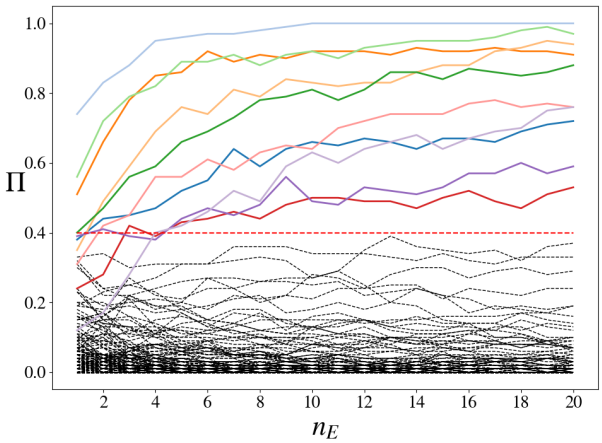
trajectories, which remain densely packed with low values. This general aspect is displayed here for one of the $N = 100$ stability selection splits. This suggests that we can isolate some, but not all, stable associations by only increasing the number of models aggregated in the r-DAA.



(a) Coefficients



(b) Stability paths using uniform ensembling

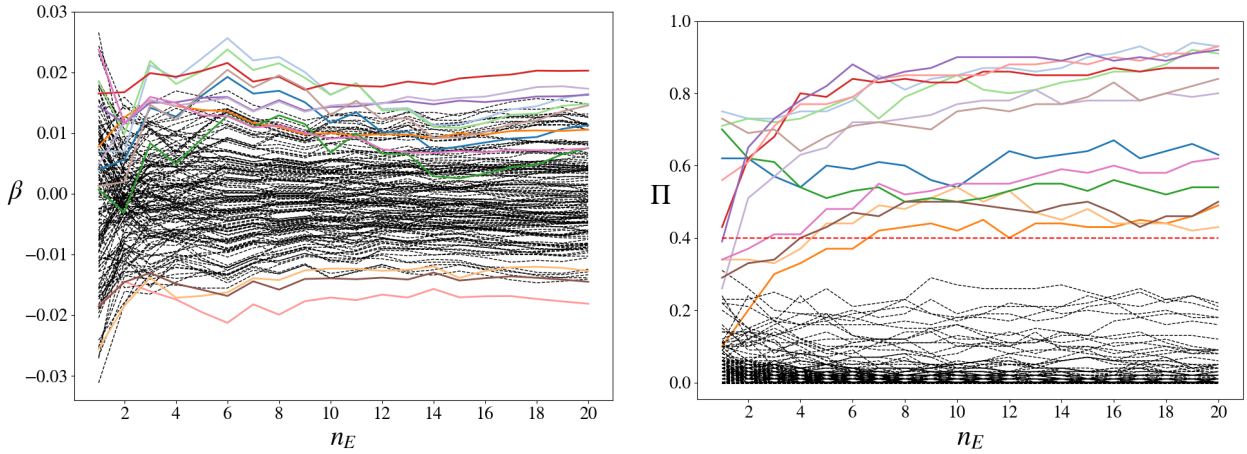


(c) Stability paths using weighted ensembling

Figure 5.8: Investigation of the ROIs associated with SRS in thickness. Each line corresponds to a ROI, dotted black ones are not selected as below the threshold $\pi_{\text{thr}} = 0.4$ when using as aggregation function a mean weighted by RSA correlations. The ROIs thus selected are coloured and the colours are consistent across the plots. The red horizontal dotted line highlights the threshold $\pi_{\text{thr}} = 0.4$. (a) Mean coefficients aggregated across models for a given split over the $N = 100$ splits, plotted against the number of models. (b) ROIs stability paths Π when using an uniformly weighted mean to aggregate the coefficients output from DAA on each model in the r-DAA, against the number of models used. (c) Same as the last one, except that the mean was weighted using RSA correlations. The latter strategy is used to select the ROIs associated in thickness with the SRS score, with $\Pi > \pi_{\text{thr}} = 0.4$.

Figure 5.8(b) presents the stability paths Π for each ROI with respect to the number of models n_E. This probability is computed as described in Eq. 4.3 and summarises the $N = 100$ stability selection splits. Here, the aggregation function f is the mean function. The results highlight that the most stable ROIs, characterised by higher probabilities Π, are easily identifiable. The stable paths reach a plateau with the number of models (i.e., the considered regularisation hyperparameter) between 5 and 10 models, depending on the considered ROI. Figure 5.8(c) is similar to Figure 5.8(b), but the procedure uses the RSA-weighted average as the aggregation function f instead of a simple average (see Supplementary Alg. C.2). The results suggest that stability is only slightly affected by the choice of the aggregation function f. We keep the RSA-weighted average because it may be more robust to outlier models. We apply a threshold of $\pi_{\text{thr}} = 0.4$ (as illustrated by the dashed red line) to retain stable associations, displayed in colours. These colours correspond to the same ROIs throughout the different plots.

Comparing Figure 5.8(b) or Figure 5.8(c) with Figure 5.8(a) reveals that a high coefficient amplitude is not always a good indicator of stability. In fact, the ROIs represented by pink or light purple paths in Figure 5.8(a) are indistinguishable from other black dotted lines (i.e., unselected ROIs). However, these ROIs convey some of the most stable associations ($\Pi \simeq 0.7$ when $n_E = 20$). Note that the variability of the SRS score is not the best captured by our models on our dataset, in that it displays the smallest significant correlation with the joint latent space representations (see Table 5.1).

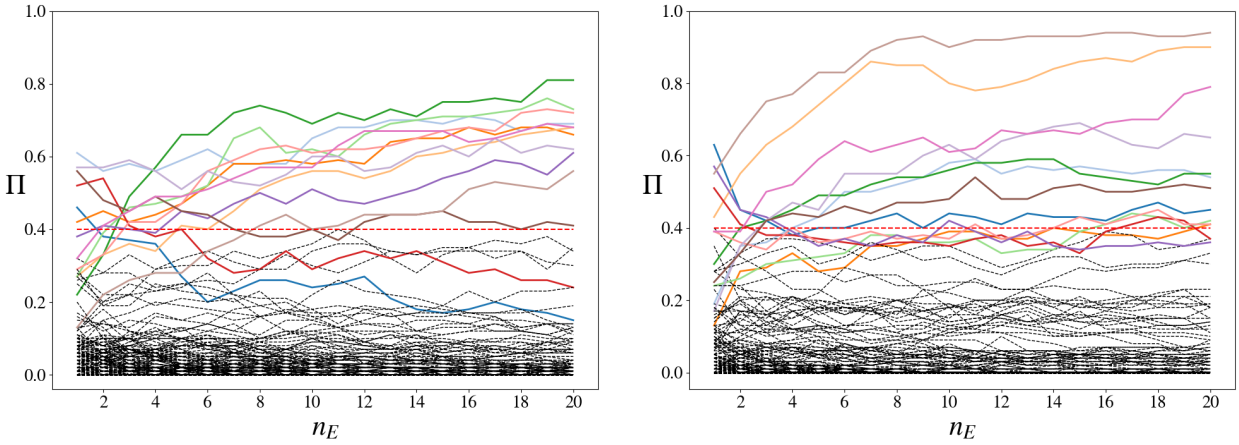


(a) Coefficients

(b) Stability paths using weighted ensembling

Figure 5.9: Investigation of the ROIs associated with SDQ-ha in area. Each line corresponds to a ROI, dotted black ones are not selected as below the threshold $\pi_{thr} = 0.4$ (highlighted by the red dotted line) when using as ensembling function a mean weighted by model ratings. The ROIs selected are coloured and the colour is consistent across the plots. (a) Aggregated coefficients with a simple mean for a given split over the $N = 100$ splits, plotted against the number of models. (b) Feature wise stability Π when using a mean weighted by the model ratings to aggregate the coefficients output from r-DAA for each model, against the number of models used.

To test the influence of the selected score and metric on the stability paths, we generate similar figures by examining the association coefficients between the SDQ-ha score and the area metric, as well as the stability paths. Very similar observations can be made, as shown in Fig. 5.9. However, this is not always the case. In Fig. 5.10, we display ROIs stability paths between SRS and area, and between SDQ-ha and thickness. The considered threshold $\pi_{\text{thr}} = 0.4$ seems less relevant for these associations. Indeed, considering Fig. 5.10(a), some ROIs (blue and red for instance) are selected because for a low number of models $n_E \in \{1, 2, 3\}$ their stability paths Π^{n_E} are above π_{thr}, but when n_E increases, their stability paths tend to decrease. Maybe these ROIs should not have been selected. On the other hand, considering Fig.5.10(b), we can see that many of the selected ROIs stability paths stay close to the threshold π_{thr}, alternating above and below it. The threshold should be studied in a more controlled environment to establish more adapted values.



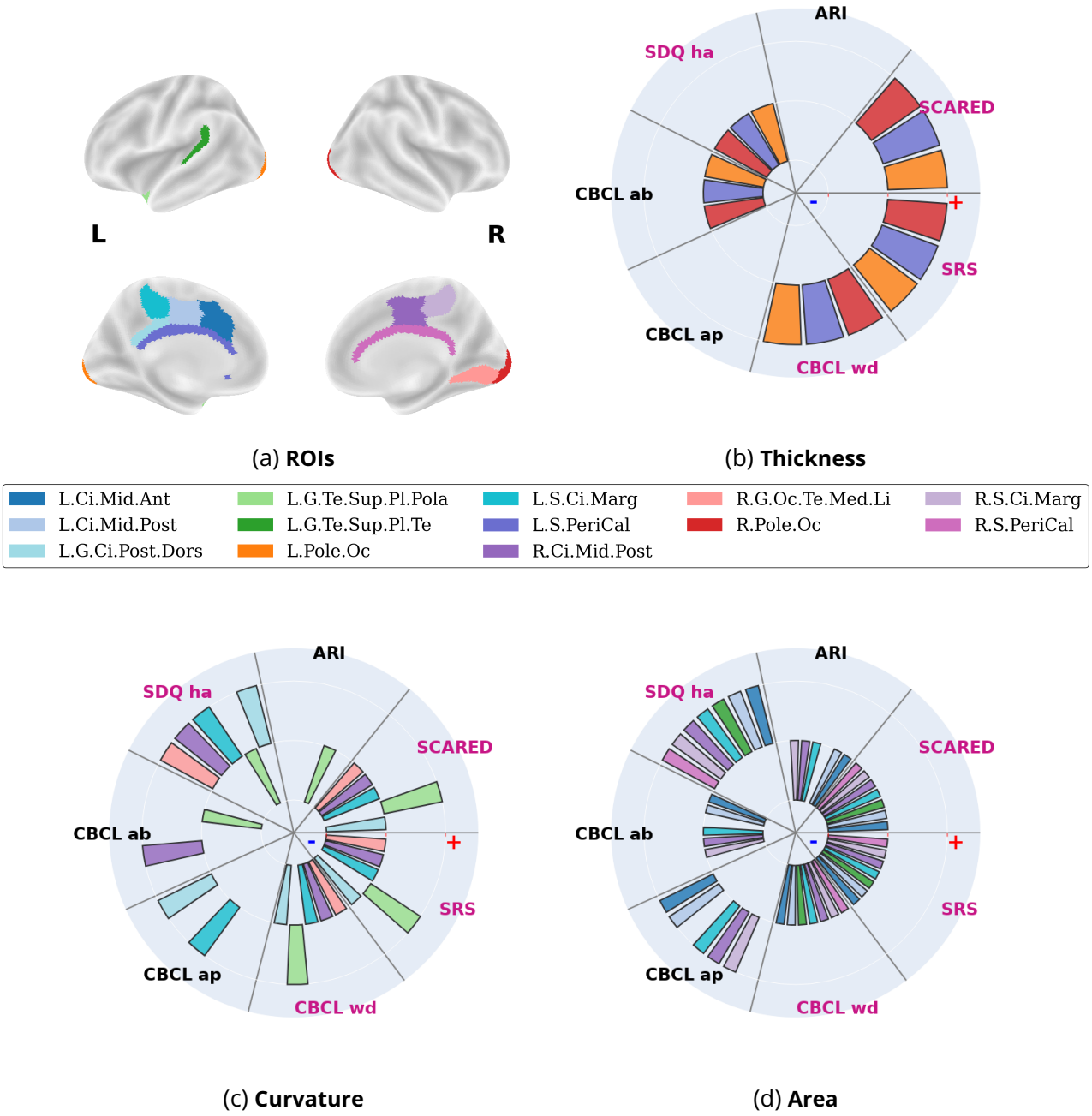
(a) Stability paths between SRS and area.

(b) Stability paths between SDQ-ha and thickness.

Figure 5.10: Investigation of the ROIs associated in more ambiguous settings. Each line corresponds to a ROI, dotted black ones are not selected as below the threshold $\pi_{thr} = 0.4$ when using as ensembling function a mean weighted by model ratings. The ROIs selected are coloured. The red horizontal dotted line highlights the threshold $\pi_{thr} = 0.4$.(a) Feature wise stability Π with SRS in area against the number of models used. (b) Feature wise stability Π with SDQ-ha in thickness against the number of models used.

5.3.4 . Transdiagnostic factor spatial support

At the core of our approach is a MoPoE-VAE network designed to construct a latent space consolidating joint information between the entire ROI view and the entire eCRF view, along with ROI- and eCRF-specific latent spaces. The interpretation brought by the Digital Avatars amounts to extracting brain-behaviour associations that effectively rely on this joint information. This strategy supports the research on the general psychopathology dimension introduced by Caspi et al. [53, 55]. In line with these ideas, our goal is to identify the brain regions and metrics that underpin these similarities. These brain-behaviour associations retained by our approach for each score-metric pair (s, m) are listed in Supplementary D. In the next paragraphs, we describe some discovered transdiagnostic associations



(a) **ROIs** (b) **Thickness**

(c) **Curvature** (d) **Area**

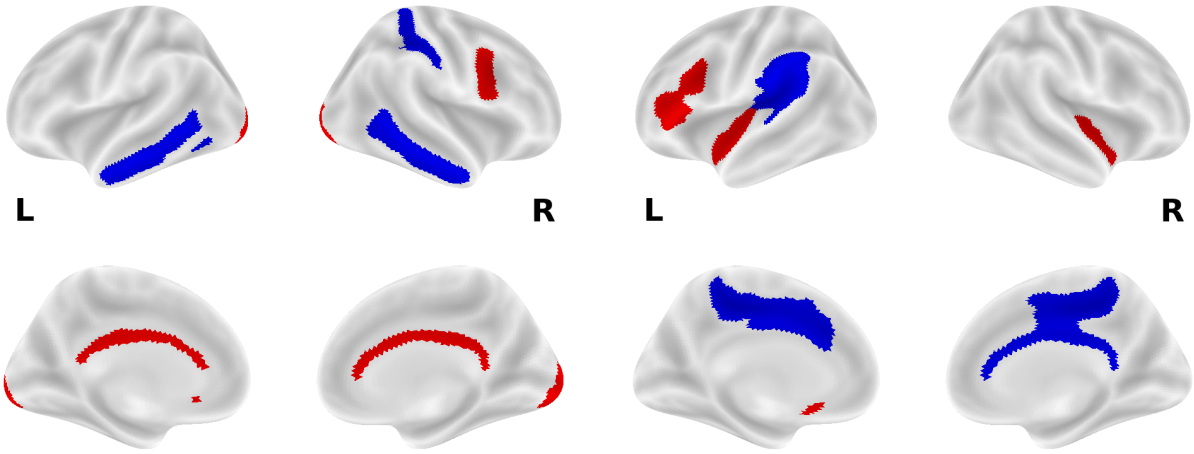
Figure 5.11: Display of ROIs associated with selected eCRF scores for a transdiagnostic assessment, e.g. SRS, SDQ-ha, SCARED and CBCL-wd, for each cortical measure. (a) ROIs colour-coded, corresponding to colour in the following polar plots. Lateral and medial view are displayed for each hemisphere (left hemisphere on the left hand side and right hemisphere on the right). The polar plots indicate the sign of each association (negative when the bar is close to the center, positive when it is close to the edge of the circle). Only retained associations are displayed. Each polar plot corresponds to a metric: (b) displays associations in thickness, (c) associations in curvature and (d) in area. L: left, R: right, Ci: cingulate, Oc: occipital, Te: temporal, S: sulcus, G: gyrus, Mid: middle, Post: posterior, Dors: dorsal, med: medial, Inf: inferior, Marg: marginal, Ant: anterior, PeriCal: pericallosal, PostCe: postcentral, Sup: superior, Li: lingual, Pl: plan, Pola: polar.

using our framework. We first describe the transdiagnostic associations that are shared across multiple disorder related dimensions, in particular ASD, ADHD and anxiety. Second, we focus on transdiagnostic associations underlying two specific disorder dimensions separately, namely ASD and ADHD.

Shared transdiagnostic factor associations. In this paragraph, we focus on transdiagnostic associations whose ROIs exhibit co-association with exactly four specific scores. We choose a limited number of scores (4 out of 7) which nevertheless represent a diversity of syndromes spanning from autism to attention disorders. These regions could be considered as components of a transdiagnostic spatial support for mental disorders. We consider the four following specific scores of interest: SRS, which assesses social interaction and is correlated with autism disorder; SCARED, which assesses fear and anxiety issues and is correlated with anxiety disorders; SDQ-ha, which is related to hyperactivity disorders; and CBCL-wd, an indicator of depression, a symptom common to most of these pathologies. For each cortical measure, Figure 5.11 highlights the ROIs found to be associated with these four scores. Among these selected transdiagnostic regions, many belong to the pericalosal and cingulate regions. These regions, whether considered with the area or curvature metrics, are consistently associated with each examined score. This means that the associations between selected ROIs and scores show similar covariations. In the identified pericalosal and cingulate regions, a decrease in both area and curvature metrics is associated with an increase in the SRS, SCARED, or CBCL-wd scores. However, the opposite is observed for the SDQ-ha score (Figure 5.11(b)-(c)). Looking at the thickness metric and grouping the curvature and area metrics, we find two disjoint sets of regions among the selected transdiagnostic regions. First, the left and right occipital poles for the thickness metric. Second, the cingulate regions for the curvature/area metrics. Only the right cingulate mid / posterior region (R.Ci.Mid.Post.) and left cingulate posterior dorsal gyrus (L.G.Ci.Post.Dors) seem to be related to the four scores for both metrics. Finally, looking at the SDQ-ha score and comparing to the grouped SRS, SCARED and CBCL-wd scores, we find that the associations identified have systematically opposite signs. Overall, the transdiagnostic regions are mostly bilateral and present a spatially smooth pattern of association with the eCRF scores.

Transdiagnostic-factor spatial support in ASD and ADHD. Recall that our method, with the presented setting, is designed to find transdiagnostic associations, in line with the general psychopathology dimension. Among the structural biomarkers involved in ASD (or ADHD), we can study the ones that can be attributed to transdiagnostic factors. In this section, we focus on transdiagnostic associations whose ROIs are specifically associated with one symptom/score. We perform this analysis for two scores. First, the SRS score, which is considered as a proxy for ASD [74]. Although the SRS score alone may not be sufficient, clinicians often use it to determine a patient's status and severity. This score has also a high correlation with diagnosis (unpublished results based on our databases) in datasets adhering to diagnosis-balanced inclusion criteria such as ABIDE I (SRS-1, 0.82), ABIDE II (SRS-1, 0.86 and SRS-2, 0.72), or EU-AIMS (SRS-2, 0.85). With the same general caveats, we use the SDQ-ha score as a proxy for ADHD-related symptoms and diagnosis [141]. The retained associations can be found in Supplementary D and are summarised below.

Markers related to SRS (ASD). In Figure 5.12, we display ROIs associated in thickness and area



(a) **Thickness** (b) **Area**

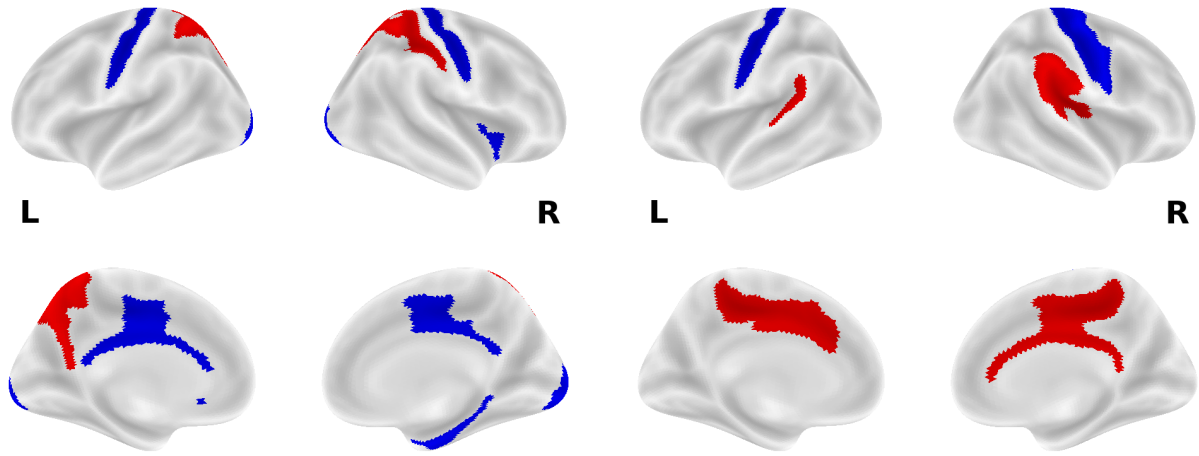
Figure 5.12: Display of ROIs associated with the SRS score in (a) thickness and (b) area. Regions coloured in red or blue display a positive or negative association respectively.

with the score SRS. Interestingly, the selected ROIs display a rather symmetrical pattern. With the thickness metric, the SRS appears to positively co-vary with the bilateral occipital poles, bilateral pericallosal region, and a right prefrontal region, while negatively co-varying with the bilateral temporal sulci and right postcentral region. With area metrics, the SRS appears to positively co-vary with the left prefrontal cortex, left subcallosal gyrus and bilateral circular inferior insular sulci, while negatively covarying with the bilateral cingulate and left parieto-temporal cortex.

Markers related to SDQ (ADHD). In Figure 5.13, we display ROIs found associated in thickness and area with the score SDQ-ha. Once again the figure displays symmetrical patterns. With the thickness metric, the SDQ-ha appears to positively co-vary notably with postcentral and superior parietal regions, including the left precuneus, while negatively co-varying with central sulci, cingulate and occipital cortices. With the area metric, the SDQ-ha appears to positively co-vary notably with bilateral parieto-temporal cortex and bilateral cingulate cortex, while negatively co-varying with bilateral central sulci and right precentral gyrus. Of note, the central sulci appear affected by the SDQ-ha score in the same direction for both area and thickness metrics.

5.4 . Discussion

In this chapter, we introduced the notion of transdiagnostic psychiatry. It relies on the hypothesis that there might exist common factors underlying psychiatric disorders, that would explain the large inter individual variability of symptom expression for a given disease, and common symptoms expressed in multiple disorders. In this context, we synthesise a few studies which consider this questions in various angles, using the HBN cohort, a population of at-risk children for developing psychiatric disorder with notable behaviour symptoms.



(a) **Thickness** (b) **Area**

Figure 5.13: Display of ROIs associated with the SDQ-ha score in (a) thickness and (b) area. Regions coloured in red or blue display a positive or negative association respectively.

The first study mostly focus on ASD. A previous work proposed an expert stratification of annotated autistic subjects based on different symptom scores, and relying mainly on the social impairment cognitive score SRS to dissociate control groups from autistic subgroups. We managed to reproduce these labels using an unsupervised integrative DL approach with a particular mVAE called sMCVAE. The sMCVAE has advantages but not modelling a joint latent space and imposing the same latent dimension for all modality-specific latent spaces are not good properties to further interpret the network.

On the other hand, we use an interpretability method dedicated to DL-based integration models using mVAEs. We demonstrate its application in the study of transdiagnostic dimensions within the HBN cohort, integrating both neuroimaging data and psychological assessments in this at-risk population with notable behavioural symptoms. Our method is endowed with a stability selection procedure to retain associations and we investigate its effects. A sufficient number of models is the only need to achieve stability. Finally, with only a prior on the expected number of associations, our method enables the identification of stable associations between measures of the brain cortical surface and symptom scores.

5.4.1 . Deep mVAE model are expressive multi-modal integration tools

In the replication study using the sMCVAE, we demonstrate the model's capacity to capture clinical related variability smoothly across a restricted number of dimensions, as well as its ability to reconstruct coherently imaging features from clinical representations, and some possibilities for interpretability. The sparsity constraint shows nice properties to control the number of dimension in the latent spaces and thus regularises the network. However, its formulation is somewhat limiting. Its training forces the modality-specific latent spaces to have the same dimension, and inherently introduces imaging variability in the clinical latent space, trough back-propagation contamination.

Therefore we would not be able to properly separate shared from modality-specific variability.

In our second study, we take advantage of the versatile definition of the latent space in the MoPoE-VAE to choose a representation setting that naturally disentangles specific and shared sources of variation between view-specific and joint latent spaces. Other work already identified the disentanglement achieved with this type of architecture [362, 222, 285], but we apply it here to neuroimaging and clinical questionnaire scores integration. We also leverage this source separation using our interpretation module based on digital avatars. By mitigating or eliminating confounding effects such as the MRI acquisition site in shared representations, we expect interpretations to be unaffected, without performing any standardisation or harmonisation of the data prior to learning. Moreover, the joint latent spaces that significantly correlate with age highlights how our models handle age information, although it is never explicitly provided. This contrasts with conventional approaches using age residualisation.

Our framework based on the MoPoE-VAE can conveniently handle incomplete data, a common issue in multi-view integration. The requirement of complete data often significantly reduces the number of subjects that can be used and impairs the statistical power of most studies. This resilience made it possible for us to use an openly available cohort with minimal missing data control and include nearly all available subjects. This has provided us with a substantial sample size collected from multiple acquisition sites. This characteristic should improve reproducibility and replicability when employing such an approach in multi-modal neuroimaging studies, aligning with state-of-the-art guidelines [203].

We also show that MoPoE-VAE networks, when equipped with a stable interpretation module, can provide associations between variables of different views. This stable interpretation module can be configured with only one a priori parameter which is the expected number of associations ($n_{\text{select}} = 12$ here).

5.4.2 . Deep mVAE models for transdiagnostic studies

Transdiagnostic theories hypothesise a common risk factor of pathologies in psychiatry, rather than focusing solely on specific diagnostic categories (such as the DSM-5 categorical diagnostic system) [125], but they are compatible with the RDoC framework which advocates multiple sets of measurements. These transdiagnostic studies consider shared risk factors such as childhood trauma (which increases the risk of all psychiatric disorders) or genetic vulnerability factors (the alleles involved often increase the risk of several psychiatric disorders at once). These approaches also stems from the observation that many psychiatric illnesses share factors, symptoms and comorbidities that cannot be easily controlled in studies which generally include population with heterogeneity. Some of these transdiagnostic works propose the existence of a general psychopathology p-factor that would act on a set of symptoms and, ultimately, disorders [53]. They typically express the global pathological variability using the p-factor, and then examine its association with brain imaging to find potential biomarkers of these common mechanisms.

The proposed MoPoE-VAE model is trained on data from an at-risk population cohort in which subjects are assessed with questionnaires, expressing symptoms of several psychiatric syndromes (such as SRS for ASD or SDQ-ha for ADHD). The learned latent representations integrate all available information into a shared and a specific latent space. The shared representations contain multivariate variability linking multiple symptom scores to imaging. The stability analysis we proposed allows the examination of this joint latent space to produce stable transdiagnostic associations between behavioural and cortical measures. Thus, when examining regions associated with multiple symptom scores in the results section, we expect to capture regions reported in the transdiagnostic literature. Our findings are discussed in this regard below:

Associations in structural MRI. We highlight fairly symmetrical regions, mostly associated in area and curvature in the cingulate regions and in thickness in the occipital poles. This is in line with many studies that report structural alterations associations with transdiagnostic factors, such as changes in cortical thickness or grey matter volume of the cingulate and occipital cortices, as part of more global patterns [140, 69, 378, 276]. Such findings have also been specifically highlighted in ASD [270, 63, 253, 106] or ADHD [12, 160, 31] studies, as well as in studies jointly examining these two conditions [299, 233]. In addition, grey matter in the occipital lobe has been specifically identified as being associated with increases in p-factor [296, 268, 297, 298].

Associations in functional MRI. functional MRI studies have also already identified the cingulo-opercular network [316] and the default mode network, supported in particular by the anterior cingulate areas [137, 238], as related to transdiagnostic factors. Other fMRI studies have identified anterior/middle cingulate and occipital areas, among others, associated with the general p-factor [108]. The same observation has been made in studies of transdiagnostic populations [115, 347], particularly with ASD and ADHD patients [46, 83, 233]. Note that we are not surprised to find the cingulate as a transdiagnostic risk marker, given its highly associative role in the brain.

Associations in diffusion MRI. Finally, dMRI studies, using fractional anisotropy of cingulum white matter tracts was also found to be associated with transdiagnostic factors [330]. Cingulum is known to be involved in cognition and emotion processing [47, 355, 93]. As these functions are often altered in the expression of psychiatric symptoms, the implication of cingulum in a common mechanism for several syndromes is very likely. Our findings support the hypothesis that cingulate and occipital regions and their related functional or structural networks are important in general psychopathology but they warrant further investigation.

5.5 . Conclusions

5.5.1 . Limitations of our transdiagnostic brain-behaviour association discovery

Results are promising, yet we would like to acknowledge a few limitations of our study and suggest possible improvements.

A smoothing effect. The DAA relies on a simplistic modelling, which surely provides highly readable results, but a more sophisticated modelling could be done. As described in Section 4.3, the associa-

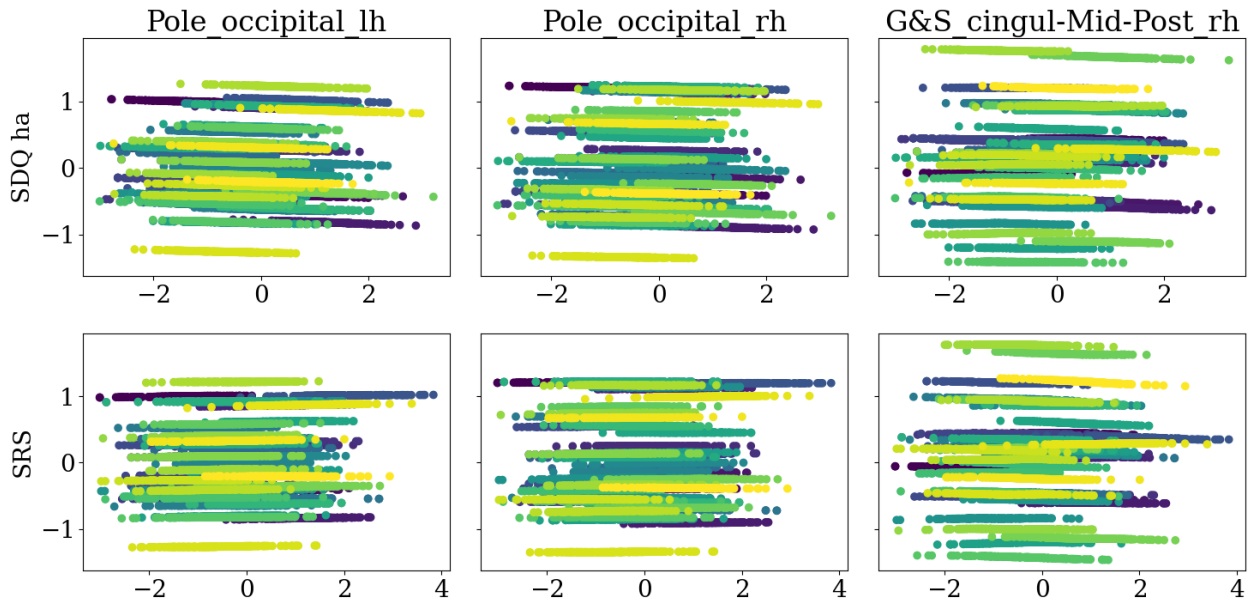


Figure 5.14: Digital avatars between cortical measures over three ROIS and two scores: SRS and SDQ-ha. The DAs displayed for occipital regions are related to their thickness, while DAs displayed for the cingulate region (right panel) represents surface area. In abscissa are represented the z-scored sampled score values (the score's name being displayed on the left of each row) while in ordinates are the z-scored values corresponding to the metric on the ROI displayed above each column.

tion between a given score and cortical measures are computed with a hierarchical modelling, considering each subject's DAs avatar as a group. Then each of these subject's association coefficients are averaged across all the L subjects from the left-out set. In cohorts such as HBN, participants present relatively low or medium symptoms scores. However, modelling these association using a linear model might be too simplistic (we could model an accelerating effect for instance), and more importantly uniformly averaging over all the participants can display a smoothing effect. This effect could be studied, in order to provide a better understanding of this modelling and finding a more optimal solution, such as a weighted average for instance, which would give more weight to participants with higher symptom scores. Some DAAs corresponding to retained associations are displayed in Fig. 5.14, in particular SRS and SDQ-ha DAs with their corresponding thickness values over bilateral occipital poles and surface area values over the middle / posterior cingulate of the right hemisphere. Each subject's DAs are represented by a line of the same colour. Each of these region were found as part of the most prominent transdiagnostic associations, yet we can barely see their effect here: SDQ-ha seems to be associated with a decreased thickness in occipital pole and decrease of area in the cingulate area. It is unclear for the SRS score. However, SDQ-ha is associated with an increase of area in this cingulate area in the association results presented in Section 5.3.4. Of course, this is displayed for a given model, trained on a given training set and DAs computed on the corresponding left-out set, which would not necessarily reflect the reality of these stable associations. But this illustrates why the effect sizes are small and could benefit from a better modelling.

Univariate associations. From our highly multivariate integration model, we derive univariate associations. We could imagine to use a different method that would be able to take into account features interactions within a block (such as correlations between clinical scores) when generating the DAs values for instance.

Computational cost. The experimental setting is quite computationally expensive. It required training 2000 models, then computing for each of these models $T = 200$ DAAs for each of the corresponding $L = 301$ subjects. This was done for every one of these models and subjects by doing 200 inferences through the corresponding model 7 times, one for each corresponding DA score value. So it resulted in roughly $840M$ inference passes through the model. Then, associations were computed using massively univariate regressions: $2000 * 301 * 7 * 444 \simeq 1.9B$ regressions between every score and imaging feature for each left-out subject's DAs of each model. Of course all these operations are highly parallelisable but they require a significant amount of computation. Moreover, storing all the corresponding results (DAs having the largest weight) demands a large storage space, a few tens of terabytes to save everything.

Overly simplistic setting. The experimental setting remains simple, for an integration problem: only two modalities are considered, other issues could arise from integrating more modalities. For instance, the training behaviour of the MoPoE-VAE, which we have not tested yet in settings with more modalities, could change, particularly if there are large discrepancies in dimensions between modalities for instance. Additionally, we did not integrate any modality with a particular geometrical structure, for which specific NNs exist to model it, such as SCNNs to learn from cortical surface over the fsaverage template (see Chapter 3). Other problems could arise from integrating a modality with such a high dimension as compared to symptoms scores, such as encoders training at different pace and introducing instability in the loss gradients.

Implementation choices. Several choices were made in the study that were not fully justified. More rigorous methodological experiments could have been led to further characterise the effect from different parameters considered in our method, such as the sampling scheme of the scores for the DAAs (for example a simple uniform or Gaussian distribution over the population's density could be used), the ensembling function f used in the r-DAA procedure (for example a max or a median function could be considered), and the number of selected associations through the decision function g, which was arbitrarily set to 12 as a good compromise. In fact, the experimental settings was different at the beginning of the experiments, and we developed a stability metric to estimate how many times an associations was replicated for different aggregation and decision functions. We made the choice of efficiency and simplicity when implementing the final experimental setting, taking advantage of our constats from the previous experiments. A proper validation would require a synthetic dataset, as a cross-validation settings is not easily conceivable here. The design of such a synthetic dataset might be very time consuming.
The stability selection could be improved as well, by testing different π_{thr} for instance, and sampling associations from different n_E models repeatedly (combinations with replacement n_E among 20) when doing the ensembling in r-DAA and selection in stability selection, in order to have better estimates of stability paths and avoid phenomenons such as presented in Fig. 5.10(a). This would also allow us to model a mean and standard deviation for each stability path.

Lack of comparison. It would be relevant to compare our method to other settings with more conventional models, such as SPLS or SGCCA, which also include a sparse selection mechanism to retain associations. However, we would need to put them in our stability selection framework in order to make them comparable. We would not be able to assess generalisability, which we do in our framework, because associations are derived from left-out subjects, i.e. not used during training our models. In the linear case, the selection mechanism directly selects features from training, without assessing how they generalise to unseen samples. It would be also great to compare our interpretation methods to more conventional approaches such as LIME of SHAP values.

5.5.2 . Perspectives

Our first contribution reported in Section 5.2.3 proposed an unsupervised replication of an expert stratification of autistic patient within the HBN cohort. The considered labels are not suited to study trandiagnostic factors, but they highlight the sMCVAE properties in capturing clinically relevant information and selecting properly relevant dimensions using the sparsity constraint. The entangled nature of the sMCVAE complicates its practical application, as it intrinsically can not display the disentanglement properties typically associated with mVAEs featuring joint and modality-specific components. But this sparsity constraint could be used to regularise a joint latent space and select the most relevant dimensions for instance. This replication study was presented during an online virtual poster session at OHBM 2021.

Our second contribution reported in Section 5.3 focuses on an integrative multi-view approach based on deep learning and featuring interpretation capacity. We discussed the main benefits of its use, its limitations, and we will now list some perspectives to improve upon of the proposed approach.

Designing a proper methodological validation setting. Implementing a synthetic benchmark or applying of stable interpretation and integrative framework in a more controlled environment might be time consuming but worth. This would reinforce the methodological aspects of this work and provide insights into how to set good hyperparameter values for every mechanism. In particular we identify the ensembling (aggregation–decision) step and selection (threshold) as key processes to be investigated further and improved upon.

Improving the disentanglement to handle site effects. We described the already reported [362, 222] disentanglement properties of explicitly modelling modality-specific and joint latent spaces. There is still some contamination of the joint latent space with unwanted confounding factors such as site effects. Improved isolation of such confounding factors in the modality-specific latent spaces where this effect originates from could be ensured by novel approaches proposing disentanglement techniques for mVAEs [89, 317].

Integrating additional views. In population imaging study, other assessments are often available. Other imaging modalities such as diffusion MRI or functional MRI could be integrated to have a more informed joint latent space, and consistency across discovered structural and functional networks could be further used to interpret the biological processes implicated in psychopathological factors.

Genotyping assessments such as relevant SNP or CNV data can be integrated as well to investigate this processes at a different scale as well.

Replication In order to validate the discovered associations using our framework, more effort should be invested in replicating our findings in different but comparable cohorts, such as the ABCD study [52].

The methodological contributions presented in Chapter 4 and exploited in this second transdiagnostic study were submitted to the journal Imaging Neuroscience and presented in a poster during the seminar IABM 2024 at Grenoble.

General conclusion

Contributions

This thesis tries to answer, to some extent, to the problematic of taking into account data structure in the learning algorithms, when analysing multimodal population cohorts for mental health. In such a highly multivariate setting, structure is mostly two-fold : each modality comes with its own structure, which is often informed by the biological processes it reflects. Then, there exist a correlation structure between the different modalities of a participant, which can be modelled as well. In order to optimally integrate multimodal data in a population study, an algorithm should be informed about these structures. Neural network offer great opportunities in this matter. Such algorithm would have the potential of uncovering biomarkers to advance mental health, in particular in psychiatry, where current treatments fall far behind the growing need for adapted and personalised clinical tools.

Modelling structure within a view. In the first part of our contributions, we focus on using structure arising from biological knowledge of the human cortex. Its continuous 3D shape interrupted by the corpus callosum makes its hemispheres homotope to a sphere. Pre-processing tools such as FreeSurfer exist to extract from brain sMRI measurements of the cortical surface and register them on a spherical template with a regular icosahedral sampling. Existing convolution operators are able to learn patterns from this spherical cortical surface representation and can be integrated in NNs. Such SCNNs can be used to solve different tasks. We identified a key application for these SCNN. Self-supervised learning has been blooming in the recent years and NNs trained using this learning paradigm offer very promising generalisation properties while not requiring human annotations. This paradigm is particularly convenient in the medical field where labelled data are often scarce and expensive. In this regard, we proposed to adapt a SSL scheme to learn from cortical surface in their spherical representations. We proposed 5 data augmentations that can be applied to such cortical surface data to allow SCNNs to learn powerful and generalisable representations using such SSL algorithm. From these 5 data augmentations, 3 are inspired from classical data augmentations applied to natural images and adapted to the cortical surface representations. The 2 other implement biologically inspired augmentations which propose to build noisy samples by replacing some values by others coming from the same participant's other hemisphere or other participant with similar cortical data. We showcase the augmentations capabilities in providing good inductive bias to the SSL, which is thus able to learn good representations. These learned representations abilities to predict 3 different phenotypes from 3 different datasets are highlighted in the experiments. These representations display improved generalisation properties as compared to their single task supervised counterparts, both to novel tasks and dataset.

Modelling structure between views. In a second part, we build up an integration framework, i.e. a model that leverages the correlation structure between views of different natures, with the capacity of providing stable associations reflecting the relationships between the different views.

We prefer to use NN based integration model, as linear models have limited capacities and can not use adapted function architectures for each view. Additionally, such models would not necessarily use participants with missing assessments and require to handle confounding factors such as imaging acquisition site effects. In particular, we identify mVAEs as suited integration models, as they naturally embed the capacity of using samples with missing views and they can use a different adapted NN architecture for each view. Moreover, mVAEs are able to separate sources of variability in view-specific and joint latent spaces, such that view-specific variability such as imaging acquisition site effects is mostly captured by the view-specific latent space. This way, no additional harmonisation technique is required.

However, mVAEs being NNs, they lack of interpretability. We design an interpretability module using individually controlled variations of a view in input space, and describe their relationship with the corresponding generated features in other views using the mVAE with hierarchical linear models. This DAA is computed using left-out subjects (i.e. not used to train the mVAE) and outputs associations between a specific view's features with every features of other views. In order to output stable associations using such NN model and interpretability module, one needs to effectively manage the variability they encounter. NNs are more subject to epistemic variability than linear models with fixed hyperparameters. We proposed a regularised version of the DAA to account for this epistemic variability. This r-DAA consists in repeating the training of the mVAE on the same training set and the interpretation on the same left-out set using the DAA, varying random initialisation and training batches. Output associations from these DAA are then ensembled using an aggregation and decision function, which outputs more robust and sparse associations. Finally, we need to control the aleatoric uncertainty due to inherent population variability to ensure that such associations are not specific to the considered sub-population. We use a stability selection procedure to retain only associations found sufficiently enough using r-DAA on multiple train / left-out splits of the population. Participants with missing assessments are included in the training sets.

This whole integration framework is then instantiated and applied to the transdiagnostic cohort HBN, composed of at-risk children presenting psychiatric symptomatic behaviours. We aim at discovering stable brain-behaviour associations in the cohort, by considering 7 questionnaires scores assessing cross-diagnostic symptoms, browsing ADHD, ASD and anxiety disorder dimensions. We compare them with cortical measures extracted from T1-weighted MRI describing ROIs of the cortex. This integration setting is in line with RDoC recommendations for studying psychiatric diseases, namely integrating multiple clinical dimensions with other assessments. Output associations can be interpreted as transdiagnostic markers, i.e. markers underlying general psychiatric disorder biological processes. Found associations display anatomical variations of cortical surface area or curvature in the cingulate cortex and thickness alterations in the primary visual cortex.

Overall, we believe that our contributions illustrate to some extent possible solutions for better using structure informed integration models in population study. In their current form, these contributions may have a limited impact in the research community. Yet, we believe that this thesis work truly opens doors with promising avenues, that we will try to sketch in what follows.

Perspectives

SCNNs trained with SSL : towards structure-informed NNs for transfer learning.

Self-supervised learning presents great generalisation properties. We saw that generalisation of NNs could be explained to some extent by inductive biases they implement. The whole paradigm of SSL is to force NNs to learn inductive bias, which could improve their generalisation capabilities. Informing the network in structure, being by using an adapted architecture the data, or by using a SSL scheme with specific augmentations designed for this type of data, can in fact improve the generalisation properties of the learned representations.
There is room for improvement over the proposed data augmentations, as we did not cross-validated every possible parameters and settings, and the proposed augmentations could be further improved and other could be proposed. For instance HemiMixUp, which relies on symmetries assumption between hemisphere, displays limited performance improvements, as compared to GroupMixUp. HemiMixUp could be improved to replace only vertices where hemispheres are the most symmetrical, thus avoiding harmfully erasing relevant asymmetry information from the learned representations. Regarding GroupMixUp, the metric used to build the groups could be improved to inject other information, such as phenotypical traits like age and sex, or clinical information. Of course these information should not be used to evaluate the representations performances afterwards. One could think of a more effective way to implement small rotations as well.
These generalisation properties embedded by the learnt representations can be leveraged to perform transfer learning. Transfer learning would allow us to train for instance a SCNN on a large cohort in general population such as UK Biobank or ABCD for a younger population and use this network for various downstream tasks in others cohorts. It could be interesting to develop deeper SCNN architectures such as ResNet or EfficientNet, featuring residual connections between layers, because transfer learning works best for more (over)parameterised architectures [286].

Integrating additional views.

As mentioned earlier, a wide variety of assessments are available in large population cohorts. Most often, each type of data is exploited separately for regression, classification and segmentation. Deep learning has been extensively used on imaging data and has demonstrated considerable benefits over traditional methods for some tasks, such as the segmentation of medical anomalies and anatomical structures [252, 162]. But in the medical field, many assessments, such as imaging or genomic data, can be more naturally represented as graphs due to their underlying biological properties (e.g. functional / structural connectomes or protein interactions). DL specific operators [40, 133] have been developed to take into account for this correlation structure when learning from such data. In our last contribution, we integrate cortical measures on ROIs as tabular data. It would be highly relevant to return to cortical spherical representation and model their actual correlation structure using dedicated convolution operators such a DiNe [394], integrated in a SCNN as presented in our first contribution. The same applies to other available modalities such as genomic data. However, this would come with several challenges as well.
The demonstrated integrative capacity of mVAEs comes from modelling modality-specific and joint latent spaces. Handling views with very different number of variables is an ongoing research question. For example, genotyping data may have millions of variables. Training a mVAE in such setting with one

view having a few dozens of variables and a genotype view with thousands, is challenging. Intuitively, learned modality specific latent spaces will have different sizes, somehow proportional to their input size, but this is not so clear for the joint representations. The sparsity constraint introduced in the sMCVAE [14] that uses variational dropout, can come in handy, if applied to the joint latent space only. However, it would require to compute a good approximation of the D_{KL} between the joint latent posterior and prior distribution, because it thus becomes not analytically tractable. This approximation is available when the joint posterior is Gaussian [259], which is not the case in the MoPoE-VAE, but it is for other mVAE such as the PoE-VAE. Additionally, there is no guarantee that these very different encoder architectures will learn at the same pace. It could be interesting to initialise these specific architectures by using SSL and specific data augmentations, such as the ones proposed in our former perspective, on an other cohort before integrating it into the mVAE.
In our transdiagnostic study, the stable associations found are characterised by rather small effect sizes. Reproducible brain-behaviour with small effect sizes have already been reported [240]. This underscores the difficulty of exploring the common processes that hypothetically contribute to the etiology of multiple psychopathologies. While individual effect sizes may be modest, the cumulative evidence supports the existence of transdiagnostic factors. One could expect an increase of the observed effect sizes by considering functional MRI data complementary to structural MRI. Association studies using fMRI offer numerous advantages, including direct measurement of brain activity and localisation of function. Current research with fMRI shows strong associations compared to structural MRI [271, 240] and would be worth investigating using our tool.

Disentanglement for harmonisation.

In the development of digital avatars, the trained NN models must be investigated to ensure they do not contain too much confounding variability, or that it is controlled, at the risk of producing biased interpretations. Importantly, the chosen mVAE model must learn disentangled representations that separate modality specific and joint variability. Indeed, confounding factors like site effect are usually specific to one modality, and such disentanglement would ensure the joint latent is devoid of such unwanted effects. Recent works propose contrastive VAE which uses deep encoders to capture higher-level semantics [5], as compared to its former linear counterpart contrastive PCA [2]. From two encoders, they typically structure the learned latent space into two parts containing the background (e.g., common to the studied population) and salient (e.g., specific to a pathology) variabilities. Followup works show that classical VAE losses alone can not effectively separate shared from salient variability, and that further constraints and regularisations are needed to satisfy the assumptions of the generative process and to promote disentanglement in the latent space [3, 66, 364, 396]. In the case of mVAEs, the same observations hold. Modelling modality-specific and joint latent spaces somehow separates joint from specific variability [362, 222], as observed in our associations discovery study. However, to ensure disentanglement, additional regularisations during training are needed [89].
Several recent works have shown that machine learning models are strongly biased by the MRI acquisition site and do not generalise well to new MRI images from sites that have never been seen before [136, 357]. While traditional residualisation techniques applied to remove the site effect marginally improves the performance of machine learning models, it does not bring any improvement for deep

learning models [103]. In this work, we show that the influence of non-interest factors, in particular the site effect, can be effectively eliminated by disentangling modality-specific and joint latent representations. Nevertheless data harmonisation remains an ongoing research area [100, 30]. For example, the OpenBHB challenge [105] on brain age prediction with site effect removal could bring new harmonisation techniques developed by the community. This resource could be used to study and validate the disentanglement properties of our mVAE. For instance, we could train a mVAE with an imaging-specific and a joint latent space between a view composed of a single preprocessing of T1-weighted MRI such as VBM and participant age as the other view. Adding disentangling constraints during the learning phase should improve the separation of imaging-specific variability and shared variability between age and imaging, thus dismissing the MRI acquisition site effects in the imaging-specific latent space. The joint representation (computed without the age for testing) could then be used to predict the age without being biased by such effects.

Replicability in other populations.

The design of a replication study will contribute to the generalisability of our findings and is a mandatory step to turn results into knowledge. In particular, it would be interesting to apply our approach to cohorts such as the the ABCD Study [52, 190], a larger population than HBN with comparable age range and comparable clinical assessments. The Duke Neurogenic Study [296, 108] is an interesting candidate as well, regrouping an undergraduate student population of about a thousand participants with diffusion and structural MRI. Multiple clinical self-reported questionnaires were completed to assess symptoms associated with psychiatric dimensions. The Dunedin Longitudinal Study [53, 297] regroups a similar sized population, but displays a longitudinal setting with numerous visits between the age of 3 and 45 years old, where the participants were scanned. Participants psychiatric symptoms were assessed during six visits between 18 and 45, where clinical questionnaire were used to assess symptoms associated with psychiatric dimensions and DSM diagnoses were set. These studies were not explicitly focused on transdiagnostic research, but their design, including comprehensive assessments with a multidisciplinary context, has provided valuable insights into the understanding of psychiatric disorders from a transdiagnostic perspective. Discovering similar transdiagnostic imaging markers in population with different age range could provide a strong evidence for the implication of these markers with general psychopathology factors.

Contributions

Peer-reviewed journal

1. **Corentin Ambroise**, Josselin Houenou, Antoine Grigis and Vincent Frouin, *Interpretable and integrative deep learning for discovering brain-behaviour associations*, 2024, submitted to Scientific Reports, preprint: https://hal.science/hal-04471394.

2. Benoit Dufumier, Antoine Grigis, Julie Victor, **Corentin Ambroise**, Vincent Frouin, Edouard Duchesnay, *OpenBHB: a Large-Scale Multi-Site Brain MRI Data-set for Age Prediction and Debiasing*, 2022, NeuroImage, https://doi.org/10.1016/j.neuroimage.2022.119637.

3. Antoine Grigis, Chloé Gomez, Jordy Tasserie, **Corentin Ambroise**, Vincent Frouin, Béchir Jarraya, Lynn Uhrig, *Predicting cortical signatures of consciousness using dynamic functional connectivity graph-convolutional neural networks*, 2020, bioRxiv, https://doi.org/10.1101/2020.05.11.078535.

Publications in international conferences

1. **Corentin Ambroise**, Antoine Grigis, Edouard Duchesnay, Vincent Frouin, *Multi-view variational autoencoders allow for interpretability leveraging digital avatars: Application to the hbn cohort*, 2023, International Symposium on Biomedical Imaging (ISBI), https://doi.org/10.1109/ISBI53787.2023.10230552.

2. **Corentin Ambroise**, Vincent Frouin, Benoit Dufumier, Edouard Duchesnay and Antoine Grigis, *MixUp brain-cortical augmentations in self-supervised learning*, 2023, Machine Learning in Clinical Neuroimaging (MLCN), https://doi.org/10.1007/978-3-031-44858-4_10.

Other communications

1. **Corentin Ambroise**, Josselin Houenou, Antoine Grigis and Vincent Frouin, *Interpretable and integrative deep learning for discovering brain-behavior associations with stability analysis*, 2024, Intelligence Articificiel pour l'Imagerie Biomédicale (IABM, french colloque)

2. **Corentin Ambroise**, Antoine Grigis, Edouard Duchesnay, Vincent Frouin, *Multi-view variational autoencoders allow for interpretability leveraging digital avatars: Application to the hbn cohort*, 2023, Intelligence Articificiel pour l'Imagerie Biomédicale (IABM, french colloque)

3. **Corentin Ambroise**, Vincent Frouin, Edouard duchesnay and Antoine Grigis, *Surfify: a geometric deep learning based Python framework for cortical surface analysis*, 2022, OHBM

4. **Corentin Ambroise**, Angeline Mihailov, Vincent Frouin and Antoine Grigis, *Multi-modal Latent Variable Model could help individuals stratification: application to HBN cohort*, 2021, Organisation for Human Brain Mapping (OHBM)

Appendices

A - Parameters for SSL on cortical surface

A.1 . Architecture and hyperparameters

Main parts	Parameters	
	name	values
SCNN encoder	DiNe channels	$128(64 \times 2) - 128 - 256 - 256$
	Activation	ReLU
	Pooling	Average [394]
	Latent dim (linear layer)	128
SimCLR [60]	MLP projector	$256 - 128$
	Temperature τ	2
	Learning rate	$2e^{-3}$
	Epochs	300 to 400
Supervised	Predictor output dim	regression: 1 / classification: 2
	Output activation	Softmax (classification)
	Learning rate	$5e^{-4}$
	Epochs	100
Optimization	Optimizer	Adam(betas=(0.9, 0.999))
	Weight decay	$1e^{-6}$
	Batch size	1024
Logistic and Ridge	L2 penalty strength	{0.01, 0.1, 1, 10, 100}

Table A.1: Description of all hyperparameters used for the different experiments. Rows with multiple values indicate that this parameter was optimized for each setting. Values separated by $-$ indicate the different output dimensions of a neural network.

A.2 . Augmentation hyperparameters

Augmentations	**Parameters**		**Probabilities** (p_{aug})
	name	values	
SurfCutOut	R	$\left[\!\left[1, g\left(\frac{P^O}{4}\right)\right]\!\right]$	0.5
SurfNoise	σ_1	$[0.1, 2]$	0.5
SurfBlur	σ_2	$[0.1, 1]$	0.5
HemiMixUp	p	0.3	0.5
GroupMixUp	p	0.4	0.5
	K	30	
	PCA_{dim}	20	

Table A.2: Description of the hyperparameters retained for each data augmentation. P is the number of vertices ($P \simeq 10k$ for a 5th order icosahedron). The function $g : x \to y$ computes the number of rings y that cuts out at most x vertices. Here $x = \frac{1}{4}$ of the total number of vertices of an O-order icosahedron.

B - Representational Similarity Analysis

Assessing model quality in an unsupervised setting remains an open issue. To investigate the learned latent representations, we derive a Representational Similarity Analysis (RSA) [206] between these representations and some measures on subjects (e.g. clinical scores or other covariates). We compute the subject-pairwise dissimilarity matrices in the latent space (modality-specific and joint) using the euclidean distance. We derive the same subject-pairwise dissimilarity matrices for the target measures. Finally, the Kendall rank correlation coefficient (Kendall τ) enables the comparison of these dissimilarity matrices emphasising the captured information. RSA is used twice in the present work. First we use it as a component of our framework to weight the contribution of a model when aggregating its extracted associations in the regularised digital avatar analysis (r-DAA) (see Section 4.4.1). Second we consider it to evaluate the information contained in the different latent representations (see Section 5.3.2).

B.1 . Model quality scoring used in r-DAA

We hypothesise that the latent representations of a good model to understand features from one modality with respect to all other should contain some amount of this modality feature-related variability. We evaluate each trained mVAE using RSA output Kendall τ between the model's joint latent space and each feature from this modality (τ_s), defined as $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{|S|}) \in [-1, 1]^{|S|}$. For each of the n_E models, we use it's average Kendall $\hat{\tau} = \frac{1}{|S|} \sum_{s=1}^{|S|} \tau_s$ across the modality's features as its quality score, used to weight the aggregation function in the ensembling of the n_E corresponding DAAs (see section 4.4.1).

B.2 . Assessing the latent representation spaces with RSA

We report in Table 5.1 aggregated results from RSAs applied to each split and corresponding models. We report the average $N \times n_E$ Kendall τ between each different measure and latent space. In the table we report the significance of these correlations. We compute it by considering the p-values of the Kendall τ statistics (corresponding to the rejection of the null hypothesis). We correct them for multiple testing using the Hommel [170] correction for multiple dependent tests and multiple independent test using Bonferroni. Then we consider the median of the corrected p-values. We chose the value of 1% for the significance threshold.

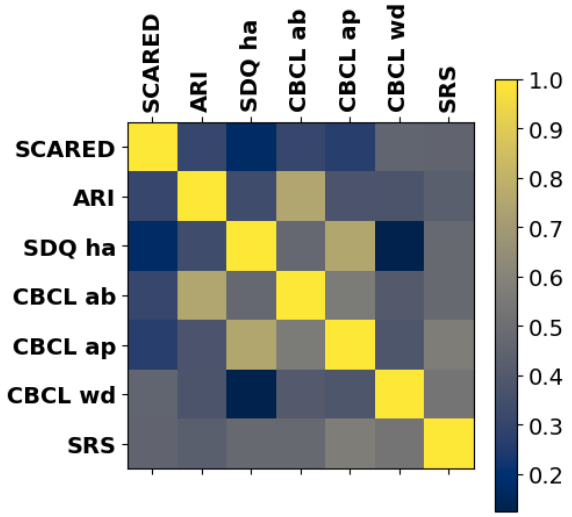


Figure B.1: Correlations between the 7 eCRF scores.

C - Algorithms for stable brain-behaviour discovery in HBN using r-DAA

We display below the pseudocode of the two procedures developed for our association discovery pipeline, relying on r-DAA and stability selection. At line 11, the quality metric is set either to 1 (quality is assumed to be uniformly correct or to the Kendall τ coefficient (see Alg. C.2).

C.1. Finding associations with a non-weighted or weighted r-DAA (DAA + ensembling)

Algorithm 1 Finding associations with a non-weighted or weighted r-DAA

Input R: a set of ROIs
M: a set of metrics
S: a set of eCRF scores
X: a list containing train/left-out subjects splits.

1: $p \leftarrow |R| * |M| * |S|$ ▷ num of features
2: $n_E \leftarrow 20$ ▷ num of models
3: $N \leftarrow 100$ ▷ num of splits
4: $n_{\text{select}} \leftarrow 12$ ▷ num of sel assocs
5: $\pi_{\text{thr}} \leftarrow 0.4$ ▷ stability path thr
6: **for** $j \in \{1, \dots, N\}$ **do**
7: $x^j_{\text{train}}, x^j_{\text{left-out}} \leftarrow X[j]$ ▷ a random split
8: **for** $k \in \{1, \dots, n_E\}$ **do**
9: Fit a $MoPoE^j_k$ model on x^j_{train} using init random weights θ^k
10: Compute the DAA association weights $\beta^j_k \in \mathbb{R}^p$ on $x^j_{\text{left-out}}$
11: Compute weights $\alpha^j_k \in \mathbb{R}_+$ assessing the model quality ▷ 1 or coefs as C.2
12: **end for**
13: $\beta^j \leftarrow [\beta^j_1, \dots, \beta^j_{n_E}]$
14: $\alpha^j \leftarrow [\alpha^j_1, \dots, \alpha^j_{n_E}]$
15: $\alpha^j \leftarrow \frac{1}{\sum_k \alpha^j_k} \alpha^j$ ▷ norm the weights
16: Select an aggregation function f
17: $f \leftarrow$ weighted mean operator ▷ we chose one for simplicity
18: Compute aggregated scores $f(\beta^j, \alpha^j) \in \mathbb{R}^p$
19: $A^j \leftarrow f(\beta^j, \alpha^j)$
20: Compute decision support $g(A^j) \in \{0, 1\}^p$ by selecting the top n_{select} associations
21: $S^j \leftarrow g(A^j)$
22: **end for**
23: Compute the selection probability of each feature $\Pi^{n_E} \in \mathbb{R}^p$
24: $\Pi^{n_E} \leftarrow \frac{1}{N} \sum_{j=1}^{N} S^j(n_E)$ ▷ function of n_E
25: Compute stable association wrt. the number of models n_E
26: $S^{\text{stable}} \leftarrow \{k : \max_{n_E \in \{1, \dots, 20\}} \Pi^{n_E} > \pi_{\text{thr}}\}$ ▷ stable associations

Output S^{stable}

(a) DAA
(b) ensembling
(c) stability selection

C.2 . Computing the model ratings to weight their associations

Algorithm 2 Computing the coefficients to weight the model assocations

Input $x^j_{\text{left-out}}$: the left-out subjects of split j
$MoPoE^j_k$: a trained model k on the x^j_{train} set
S: a set of eCRF scores

$L \leftarrow |x^j_{\text{left-out}}|$ ▷ num of left-out subj
$z_k \leftarrow MoPoE^j_k(x^j_{\text{left-out}}) \in \mathbb{R}^{L \times d}$ ▷ joint latent d-dim reps
Compute the pairwise joint latent dissimilarity matrix $C^j_k \in \mathbb{R}^{L \times L}$
$C^j_k \leftarrow \left[\|z^m_k - z^n_k\|^2_2 \text{ for } (m,n) \in \{1,\dots,L\}^2\right] \in \mathbb{R}^{L \times L}$ ▷ pairwise l_2 dist
for $s \in S$ **do**
Compute the pairwise eCRF score dissimilarity matrix $S^s \in \mathbb{R}^{L \times L}$
$S^s \leftarrow \left[\|s^m - s^n\|^2_2 \text{ for } (m,n) \in \{1,\dots,L\}^2\right] \in \mathbb{R}^{L \times L}$
Compute the associated Kendall $\tau^{j,s}_k \in \mathbb{R}$ between C^j_k and S^s
end for
Average the obtained Kendall $\tau^{j,s}_k \in \mathbb{R}$ across the eCRF scores
$\tau^j_k \leftarrow \frac{1}{|S|} \sum_{s \in S} \tau^{j,s}_k$

Output τ^j_k

D - All associations

Score	SRS	SCARED	ARI	SDQ ha
Metric				
Thickness	L.G.Te.Mid L.Pole.Oc L.S.PeriCal L.S.Te.Inf R.G.Te.Mid R.Pole.Oc R.S.PeriCal R.S.PostCe R.S.PreCe.Inf. R.S.Te.Inf	L.Ci.Mid.Post L.G.Ci.Post.Dors L.G.Pa.Sup L.G.Te.Mid L.Pole.Oc L.S.PeriCal L.S.Te.Inf R.G.Oc.Te.Med.PH R.G.Pa.Sup R.G.Te.Mid R.Pole.Oc R.S.Inter.Prim.Jens R.S.PeriCal R.S.PreCe.Inf. R.S.Te.Inf	L.FrMarg L.Ci.Mid.Post L.G.Oc.Te.Med.PH L.G.PostCe L.G.Te.Inf R.G.Oc.Te.Med.PH R.G.PostCe R.G.Te.Inf R.Lat.Fis.Ant.Vert R.Pole.Te R.S.Fr.Inf R.S.Fr.Mid	L.Ci.Mid.Post L.G.Pa.Sup L.G.Precu L.Pole.Oc L.S.Ce L.S.PeriCal R.Ci.Mid.Post R.G.Ci.Post.Dors R.G.Ins.St R.G.Oc.Te.Med.PH R.G.Pa.Sup R.Pole.Oc R.S.Ce R.S.PostCe
Meancurv	L.Ci.Mid.Post L.G.Ci.Post.Dors L.G.Ci.Post.Ventr L.G.Te.Sup.Pl.Pola L.S.Ci.Marg L.S.PeriCal R.Oc.Inf R.Ci.Mid.Post R.G.Ci.Post.Ventr R.G.Oc.Te.Med.Li R.G.Te.Sup.Pl.Pola R.Pole.Oc R.S.Circ.Ins.Sup R.S.PeriCal	L.Ci.Mid.Post L.G.Ci.Post.Dors L.G.Ci.Post.Ventr L.G.Te.Sup.Pl.Pola L.S.Ci.Marg L.S.Pa.Oc L.S.PeriCal R.Oc.Inf R.Ci.Mid.Post R.G.Ci.Post.Ventr R.G.Oc.Te.Med.Li R.Pole.Oc R.S.Circ.Ins.Sup R.S.PeriCal	L.G.Ins.St L.G.SubCal L.G.Te.Sup.Pl.Pola L.S.Oc.Te.Med.Ling L.S.Pa.Oc L.S.PreCe.Sup. R.Transv.FrPole R.G.PostCe R.G.Te.Sup.Pl.Pola R.S.Ce R.S.PeriCal	L.G.Ci.Post.Dors L.G.Ins.Lg.S.Cent.Ins L.G.Te.Sup.Pl.Pola L.S.Ci.Marg L.S.Pa.Oc L.S.SubPa R.Oc.Inf R.Ci.Mid.Post R.G.Oc.Te.Med.Li R.G.Pa.Inf.Ang R.G.Pa.Inf.Supr
Area	L.Ci.Mid.Ant L.Ci.Mid.Post L.G.Fr.Inf.Tri L.G.Pa.Inf.Supr L.G.SubCal L.G.Te.Sup.Pl.Te L.Lat.Fis.Post	L.ParaCe L.Ci.Mid.Ant L.Ci.Mid.Post L.G.Te.Sup.Pl.Te L.Lat.Fis.Post L.S.Ci.Marg L.S.Orb.Med.Olfact	L.Ci.Mid.Ant L.Ci.Mid.Post L.G.PostCe L.G.SubCal L.S.Ce L.S.Ci.Marg L.S.Te.Sup	L.Ci.Mid.Ant L.Ci.Mid.Post L.G.Te.Sup.Pl.Te L.S.Ce L.S.Ci.Marg R.Ci.Mid.Post R.G.Pa.Inf.Supr

	L.S.Ci.Marg	R.Ci.Mid.Post	R.Ci.Mid.Post	R.G.PreCe
	L.S.Circ.Ins.Inf	R.G.Rect	R.G.PreCe	R.G.Te.Sup.Pl.Te
	L.S.Fr.Inf	R.G.Te.Sup.Pl.Te	R.S.Ce	R.Lat.Fis.Post
	R.Ci.Mid.Post	R.S.Ci.Marg	R.S.Ci.Marg	R.S.Ce
	R.S.Ci.Marg	R.S.Orb.Med.Olfact		R.S.Ci.Marg
	R.S.Circ.Ins.Inf	R.S.PeriCal		R.S.PeriCal
	R.S.PeriCal			

Score	CBCL ab	CBCL ap	CBCL wd
Metric			
Thickness	L.G.Oc.Te.Med.PH	L.FrMarg	L.Ci.Mid.Post
	L.G.PostCe	L.Ci.Mid.Post	L.G.Pa.Sup
	L.G.Te.Inf	L.G.Oc.Te.Med.PH	L.Pole.Oc
	L.Pole.Oc	L.G.Te.Inf	L.S.PeriCal
	L.S.Fr.Inf	L.S.Fr.Mid	L.S.PreCe.Sup.
	L.S.PeriCal	R.Ci.Mid.Post	L.S.Te.Inf
	L.S.Te.Inf	R.G.Ci.Post.Dors	R.Ci.Mid.Post
	R.G.Oc.Te.Med.PH	R.G.Ins.St	R.Pole.Oc
	R.G.PostCe	R.G.Oc.Te.Med.PH	R.S.Ce
	R.G.Te.Inf	R.G.Te.Inf	R.S.PostCe
	R.G.Te.Mid	R.Pole.Te	R.S.PreCe.Inf.
	R.Lat.Fis.Ant.Vert	R.S.Ce	
	R.Pole.Oc	R.S.Fr.Mid	
Meancurv	L.Ci.Mid.Post	L.G.Ci.Post.Dors	L.ParaCe
	L.G.SubCal	L.G.Ins.Lg.S.Cent.Ins	L.G.Ci.Post.Dors
	L.G.Te.Sup.Pl.Pola	L.G.SubCal	L.G.Ins.Lg.S.Cent.Ins
	L.S.Pa.Oc	L.S.Ci.Marg	L.G.Te.Sup.Pl.Pola
	L.S.PeriCal	L.S.Pa.Oc	L.S.Ci.Marg
	L.S.PreCe.Sup.	L.S.SubPa	L.S.Pa.Oc
	R.ParaCe	R.G.Oc.Te.Med.PH	R.Ci.Mid.Post
	R.Ci.Mid.Post	R.G.Pa.Inf.Ang	R.G.Oc.Te.Med.Li
	R.G.Te.Sup.Pl.Pola	R.G.Pa.Inf.Supr	R.G.Pa.Inf.Ang
	R.S.Circ.Ins.Sup	R.G.PostCe	R.Pole.Oc
	R.S.PeriCal	R.S.Ce	R.S.Ci.Marg
		R.S.Fr.Sup	R.S.PeriCal
Area	L.Ci.Mid.Ant	L.Ci.Mid.Ant	L.ParaCe
	L.Ci.Mid.Post	L.Ci.Mid.Post	L.Ci.Mid.Ant
	L.G.Fr.Inf.Tri	L.S.Ce	L.Ci.Mid.Post

L.G.PostCe	L.S.Ci.Marg	L.G.Fr.Inf.Tri
L.G.SubCal	L.S.Te.Sup	L.G.Te.Sup.Pl.Te
L.S.Ci.Marg	R.Ci.Mid.Post	L.Lat.Fis.Post
L.S.Fr.Inf	R.G.Fr.Mid	L.S.Ci.Marg
L.S.Te.Sup	R.G.Oc.Te.Med.PH	L.S.Circ.Ins.Inf
R.Ci.Mid.Post	R.G.PreCe	L.S.Fr.Inf
R.S.Ci.Marg	R.G.Te.Sup.Pl.Te	R.Ci.Mid.Post
R.S.Circ.Ins.Inf	R.S.Ce	R.G.Pa.Inf.Supr
	R.S.Ci.Marg	R.G.PreCe
		R.G.Te.Sup.Pl.Te
		R.Lat.Fis.Post
		R.S.Ci.Marg
		R.S.PeriCal

Table D.1: Retained associations for each score and metric. Blue indicates a negative association and red denote a positive one. L: left, R: right, S: sulcus, G: gyrus, Lat: lateral, Ci: cingul, Pa: parietal, Ce: central, Oc: occipital, Te: temporal, Fr: front, Orb: orbital, Ins: insula, Post: posterior, Mid: middle, Ant: anterior, Med: medial, Sup: superior, Inf: inferior, Ventr: ventral, Dors: dorsal, PostCe: postcentral, PreCe: precentral, ParaCe: paracentral, PeriCal: pericallosal, Marg: marginal, Pl: plan, Tri: triangular, Circ: circular, Supr: supramarginal, Ang: angular, PH: parahippocampal, SubPa: subparietal, Transv: transverse, FrPole: frontopolar, Pola: polar, SubCal: subcallosal, St: short, Fis: fissure, Olfact: olfactory, Rect: rectus, Precu: precuneus, FrMag: fronto-marginal, FrMarg: frontomargin, Inter: intermedius, Li: lingual, Jens: Jensen, Vert: vertical, Prim: primus.

Résumé en Français

Introduction

La santé mentale est un enjeu majeur puisque 970 millions de personnes vivaient avec un trouble mental en 2019, selon l'Organisation Mondiale de la Santé (OMS)[1]. Cependant, ces maladies sont souvent négligées et les ressources investies dans leur traitement sont insuffisantes. Cela s'explique en partie par le fait qu'elles sont assez mal comprises. Le développement récent de systèmes non invasifs (notamment l'IRM) et à haut débit (notamment les séquenceurs ADN) permettant d'obtenir des informations sur les systèmes biologiques et leur fonctionnement pourrait changer cette situation. Mais d'autres paradigmes devraient changer pour permettre une progression globale des connaissances en matière de pathologie cérébrale. Pour les troubles psychiatriques, l'approche conventionnelle consiste actuellement essentiellement à considérer un diagnostic ou une condition unique, et à utiliser un modèle statistique pour les prédire à l'aide d'une mesure biologique prise isolément telle qu'une modalité d'imagerie. Cependant, il a été reconnu que les résultats obtenus avec cette stratégie ne sont pas satisfaisants et qu'elle limite la progression des connaissances en psychiatrie. Les raisons de cet insuccès pourraient tenir à l'hétérogénéité reconnue des conditions et diagnostics définis par des outils de classification tels que le Diagnostic and Statistical Manual of Mental Disorders (DSM). Le phénomène de comorbidité entre les pathologies et les interactions entre les processus biologiques, qui existent à différents niveaux, sont négligés par les études basées sur le diagnostic DSM. D'autres initiatives, telles que le Research Domain Criteria (RDoC), préconisent une approche plus globale de l'étude des troubles psychiatriques en incorporant diverses données qui reflètent différents niveaux de la complexité organisationnelle de la vie (par exemple, l'imagerie, la génétique et les symptômes). Les principes du RDoC suggèrent qu'une description approfondie d'une pathologie nécessite la prise en compte de multiples dimensions qui peuvent être partagées entre différents syndromes psychiatriques et peuvent même contribuer à la variabilité non pathologique.

En utilisant des données qui suivent ces recommandations, ce travail de thèse vise à fournir des outils compatibles avec les critères RDoC, et en particulier des outils reposant sur les réseaux de neurones artificiels. Ces outils doivent permettre de modéliser au mieux les informations de structure issues de connaissances a priori des composants biologiques disponibles, dans l'imagerie de population destinée à l'étude des troubles psychiatriques.

Contexte

L'Imagerie par Résonance Magnétque (IRM) permet d'étudier les tissus du cerveau et son activité. En particulier, l'IRM permet d'inspecter précisément la structure du cerveau et en particulier du cortex, qui constitue sa couche la plus externe. Il représente la majeure partie de la matière grise du

[1] Santé mentale de l'OMS topic.

cerveau, c'est-à-dire les tissus regroupant la plupart des cellules neuronales. L'activité des neurones se présente sous la forme d'un signal électrique et chimique voyageant à travers le cerveau et entre deux zones du cortex ou entre le cortex, le cervelet et les composants sous-corticaux. L'étude globale du cerveau offert par l'IRM peut fournir des indications précieuses pour la compréhension des fonctions et les dysfonctionnements cérébraux. La structure très plissée du cortex lui permet d'augmenter considérablement sa surface et son volume de matière grise, mais rend son étude difficile en raison de sa forme particulière. Des logiciels experts tels que FreeSurfer [86, 120] permettent d'extraire des mesures géométriques de ce tissu et de les représenter sur un modèle commun. Celui-ci permet de mieux préserver les propriétés topologiques du cortex que ce que ne pourraient faire des images 3D.

Pour étudier les caractéristiques et les troubles du cerveau, des consortiums recueillent de grandes quantités de données afin de mettre en évidence des biomarqueurs d'intérêt en recrutant des sujets en population générale. Cette approche permet également de mieux prendre en compte les caractéristiques spécifiques de chaque population. Aujourd'hui, de vastes projets de recherche ont permis de créer des ressources partagées et ouvertes à partir des données de santé d'une large population. Parmi celles-ci, les cohortes d'imagerie de population, promeuvent un phénotypage de haute qualité grâce à l'imagerie et à l'évaluation psychologique, tout en recrutant spécifiquement un très grand nombre de personnes dans le cadre d'un protocole de recherche. Ces initiatives sont conformes au besoin croissant de données pour les modèles statistiques récents tels que l'apprentissage profond. Il a également été récemment démontré que la découverte d'associations reproductibles entre le cerveau et des phénotypes comportementaux complexes nécessite des milliers d'individus [240]. Des problèmes spécifiques apparaissent dans ces cohortes, la collecte d'un vaste ensemble de données nécessite le regroupement de données provenant de sources multiples et de différents sites d'acquisition d'imagerie. Cela pose de nouveaux défis, tels que l'harmonisation des effets de site, dont on sait qu'elle nuit à la qualité de la modélisation en aval [136, 357].

Une grande variété d'outils (par exemple IRM et microarrays) sont disponibles pour mieux comprendre la structure et le fonctionnement du cerveau. Les chercheurs ont commencé à trouver des biomarqueurs du cerveau responsables des fonctions cognitives ou des pathologies cérébrales. Certains marqueurs pathologiques peuvent être spécifiés par un expert humain à partir d'images : une tumeur cérébrale est souvent identifiable par un radiologue. D'autres sont beaucoup plus subtils, et nous ne pouvons pas les identifier en regardant simplement un scan IRM (par exemple, les caractéristiques radiomiques subtiles qui varient pour qualifier les cancers). Pour aider les chercheurs en imagerie médicale à comprendre les mécanismes sous-jacents des pathologies ou des fonctions, les ingénieurs se sont appropriés les algorithmes d'apprentissage basés sur les données, qui peuvent être directement appliqués pour trouver des biomarqueurs. Ces approches tendent à modéliser les mécanismes sous-jacents qui régissent les données, afin de pouvoir relier correctement les phénotypes observés à la biologie (par exemple, la conversion de la maladie d'Alzheimer avec l'accélération du rétrécissement de la matière grise). L'apprentissage automatique peut-être supervisé, c'est-à-dire guidé par des annotations humaines qui peuvent être très coûteuses, ou non-supervisé, auquel cas l'objectif de l'algorithme d'apprentissage est plutôt de modéliser la

distribution des données.

Une sous-catégorie majeure des algorithmes d'apprentissage automatique, appelée apprentissage profond, regroupe tous les algorithmes d'apprentissage utilisant des réseaux de neurones artificiels. Ils ont été introduits au début des années 60 avec le Perceptron multi-couches (MLP) de Rosenblatt.[302]. Un MLP est une succession de projections linéaires, appelées couches denses, entrecoupées de fonctions d'activation non linéaires. Aujourd'hui, les réseaux de neurones sont entraînés à l'aide d'algorithme de descente de gradient [293, 10] ou leurs extensions [195] par rétropropagation [227, 366]. Cependant, en raison de la complexité des architectures mises en oeuvre, les raisons des performances des réseaux de neurones sont mal comprises. Un phénomène très surprenant est leur étonnante propriété de généralisation, qui évolue avec leur complexité, contrairement aux autres modèles d'apprentissage automatique. Par conséquent, les propriétés d'expressivité des réseaux de neurones ont été décrites pour mieux comprendre leurs propriétés de généralisation. Il semblerait que des architectures adaptées aux données, telles que les opérateurs de convolution, mettent en œuvre des biais inductifs qui pourraient être une source d'explication de ces propriétés de généralisation. Outre leurs caractéristiques de généralisation surprenantes, l'interprétation des réseaux de neurones n'est pas simple en pratique. Nous essayons de décrire quelques méthodes d'explication qui tentent de surmonter ce problème. Finalement, nous présentons un concept non supervisé qui a gagné un intérêt croissant avec les réseaux de neurones, appelé apprentissage auto-supervisé. Il est fondé sur des représentations d'apprentissage qui appliquent des biais inductifs aux réseaux en exploitant des manipulations de données soigneusement conçues.

Contributions

Modéliser la structure de la surface corticale

La modélisation de la surface corticale sous la forme d'un maillage icosaédrique est un moyen très complet d'intégrer la topologie spécifique de la surface corticale du cortex dans les données. Elle tient compte du fait que le cortex est ininterrompu chez les humains sains et que les gyri voisins sont en fait topologiquement éloignés. Le fait d'imposer une telle structure dans les données présente de nombreux avantages. Cela peut être utilisée pour réduire la dimensionnalité des données, par rapport aux images 3D produites par l'imageur, en se concentrant sur un signal intéressant pour l'étude des maladies psychiatriques. De plus, cette représentation est plus précise tout en restant spatialement cohérente par rapport aux mesures moyennes sur les régions d'intérêt. En outre, le fait de disposer d'une représentation des données structurée par des considérations biologiques peut permettre à la modélisation ultérieure de fournir des biomarqueurs plus pertinents et plus cohérents, améliorant ainsi l'interprétabilité.

Dans ce contexte, nous avons étudié les adaptations spécifiques nécessaires dédiées aux réseaux de neurones artificiels pour construire des représentations de données corticales prenant en compte la structure topologique du cortex. Nous nous appuyons sur des opérateurs de convolution adaptés existants, tels que DiNe, qui nous permettent de concevoir des réseaux de neurones convolutionels de graphe (SCNN) capables d'apprendre à partir de données de surface corticale icosaédrique.

Nous introduisons des augmentations de la surface corticale conçues pour entraîner un réseaux de neurones convolutionels de graphe dans une configuration d'apprentissage auto-supervisé (SimCLR-SCNN). Parmis les 5 augmentations de données proposées, 3 reposent sur une adaptation de ce qui est réalisé pour des images naturelles et 2 sont inspirées par des hypothèses biologiques. Elles consistent à remplacer des valeurs par d'autres réalistes, dans un cas considérant la symétrie inter hémisphérique (HemiMixUp), et dans l'autre des données de surfaces corticales d'autre individus similaires (GroupMixUp). Le réseau SimCLR-SCNN entraîné à l'aide de ces augmentations démontre sa capacité à générer des représentations dotées de fortes propriétés de généralisation. En fait, les représentations apprises à partir de données collectées dans différentes études offrent des performances prometteuses, parfois même supérieures aux approches supervisées. En particulier, l'augmentation GroupMixUp montre un potentiel pour l'apprentissage de représentations stables à travers différentes cohortes.

Au final, nous avons mis au point une méthodologie permettant d'apprendre des représentations généralisables des données de surface corticale. Elle pourrait être exploitée pour mieux initialiser les réseaux de neurones sur de grandes cohortes de population de sujets sains, avant de les transférer pour réaliser des tâches d'intérêt sur des cohortes cliniques plus petites. Nous contribuons à cet objectif important en proposant quelques augmentations de données de surface corticale, clés de voûte pour s'adapter aux nouveaux cadres d'apprentissage auto-supervisés, et les outils environnants nécessaires pour les utiliser. Notamment, le module surfify a été présenté lors de OHBM 2022 à Glasgow au cours de sessions de posters. En outre, les expériences liées à l'apprentissage auto-supervisé sur la surface corticale ont conduit à une publication pour la conférence satellite Machine Learning in Clinical Neuroimaging (MLCN) de MICCAI 2023.

Cette méthodologie répond à la question de la défintion d'approches d'apprentissage profond à partir d'IRM cérébrales pondérées T1, qui soient capables de tirer parti de la structure corticale dans ces données. Dans la partie suivante, au lieu de travailler sur des structures intra-modalité, nous proposons des approches compatibles avec l'apprentissage profond pour modéliser les relations entre diverses sources de données. Ces approches doivent fonctionner avec des types de données disponibles dans les cohortes de population, telles que l'IRM pondérée en T1 avec d'autres modalités d'imagerie, des données de génotypage ou des questionnaires cliniques.

Modéliser les structures de corrélation entre les modalités pour la psychiatrie

Aujourd'hui, la psychiatrie, tant dans sa dimension diagnostique que thérapeutique, passe d'un paradigme fondé sur l'étude des syndromes à un nouveau paradigme fondé sur la compréhension des mécanismes neurobiologiques qui les sous-tendent. La réalisation de cet objectif est un processus continu qui nécessite la combinaison d'avancées scientifiques, d'innovations technologiques et d'une approche centrée sur le patient [368]. L'identification des relations entre les comportements et les mesures cérébrales est un aspect essentiel de ce paradigme. Cependant, les comportements sont complexes et résultent souvent d'une combinaison de facteurs génétiques, environnementaux et psychologiques. En outre, le phénomène de comorbidité fréquemment évoqué en psychiatrie signifie que de multiples comportements associés à divers troubles mentaux peuvent se manifester

simultanément, et qu'un même trouble peut se manifester par des comportements différents chez des individus différents. Cette complexité remet en cause l'idée qu'une signature comportementale unique peut correspondre à un seul trouble. Comme nous l'avons introduit précédemment, l'initiative RDoC fournit des recommandations pour aborder correctement cette complexité. Contrairement aux approches traditionnelles qui visent à établir un diagnostic à partir d'un score ou d'une modalité spécifique (comme indiqué dans le DSMs), le RDoC promeut des approches dimensionnelles et transdiagnostiques. Ces dimensions sont recherchées à travers les facteurs génétiques, biologiques, environnementaux et de mode de vie dans la recherche sur la psychiatrie personnalisée.

Dans ce contexte, nous avons besoin de modèles statistiques capables de décrire ces interactions entre les différentes données disponibles. Par la suite, nous décrivons les propriétés que ces algorithmes doivent présenter pour être applicable dans notre contexte. Nous présentons les méthodes que nous avons identifiées ainsi que les méthodes proposées pour répondre à cette problématique. Ensuite, nous appliquons certaines de ces méthodes à la cohort Healthy Brain Network (HBN) pour une analyse transdiagnostique. La cohorte HBN, qui regroupe une population d'enfants à risque pour différents troubles psychiatriques, est adaptée pour mener cette étude.

Méthodes. Afin de traiter la quantité et la diversité croissante des données, et de modéliser la part de leur variabilité imputable à la maladie, nous avons besoin d'outils d'apprentissage automatique dotés des propriétés suivantes :

- **Non-supervisé ou semi-supervisé** : l'algorithme d'apprentissage ne doit pas être uniquement orienté par une prédiction diagnostique, qui ne rend pas suffisamment compte de la variabilité liée à la pathologie et limite la quantité de données disponibles. Envisager un cadre semi-supervisé, en intégrant les symptômes au lieu du diagnostic, permet de prendre en compte à la fois les sujets présentant des évaluations liées aux symptômes et les sujets qui en sont exempts.

- **Intégratif** : capable de traiter une quantité arbitraire de données sous diverses formes et de modéliser leur variabilité partagée, tout en intégrant la structure de corrélation intra-modalité. Par exemple, on cherche à analyser conjointement des mesures sur le cortex obtenues par IRM et des données tabulaires du dossier patient en vue de trouver des relations avec la maladie.

- **Capable de traiter les données manquantes** : toutes les cohortes multimodales ont des données manquantes, qui peuvent apparaître à différents niveaux, soit de manière sporadique dans un type d'évaluation, comme un questionnaire clinique non rempli, soit une modalité complètement manquante comme une IRM, si le patient n'a pas pu être scanné pour une raison ou une autre. De telles absences sont très fréquentes dans les études multimodales, et le modèle d'apprentissage que nous choisissons doit être capable d'utiliser des observations avec des données manquantes, au risque de voir sa puissance statistique fortement diminuée.

- **Interprétable** : nous devrions être en mesure de comprendre et d'expliquer ce que le modèle a appris. L'objectif ultime serait de décrire les mécanismes biologiques responsables des

maladies psychiatriques, découverts par le modèle. Cela est impossible si le modèle n'est pas interprétable d'une manière ou d'une autre.

Nous partons des modèles d'intégration candidats qui implémentent des approches linéaires, notamment Canonical Correlation Analysis (CCA), Partial Least Squares (PLS) et leurs extensions. Toutefois, leur incapacité à utiliser des observations avec des mesures manquantes est très limitante. Les Probabilistic Graphical Model (PGM) profonds, et en particulier les Auto-Encodeurs Variationels multi-vues (mVAE) présentent des propriétés très attrayantes. Ils sont capables d'exploiter les observations avec des données manquantes et de modéliser efficacement la structure de corrélation intra (avec des réseaux spécifiques aux modalités et des espaces latents privés) et inter-modalités (avec un espace latent commun). En outre, il a été démontré qu'ils intègrent des propriétés de désenchevêtrement, qui peuvent être utilisées pour distinguer la variabilité spécifique de la variabilité partagée. Cela peut avoir des implications importantes, car la plupart des biais ne sont pas contenus dans toutes les modalités : l'effet du site d'acquisition de l'image est un très bon exemple de ce type de biais, qui sera limité aux modalités d'imagerie. Les effets liés au sexe pourraient être limités aux modalités d'imagerie et de génétique. Les questionnaires cliniques sont conçus pour ne pas être biaisés par de tels facteurs.

Ces mVAEs ne sont en revanche pas directement interprétables, contrairement aux modèles d'intégration linéaires. Nous examinons donc les possibilités d'interprétation qui s'offrent pour les modèles présentés précédemment et proposons une méthode d'interprétation utilisant des avatars numériques (DAA) qui interprètent ce qu'un modèle a appris. Cette interprétation est obtenue en modifiant légèrement les caractéristiques d'une modalité dans l'espace d'entrée et en interrogeant le mVAE pour générer les changements correspondants dans les autres modalités. Ce module d'interprétation produit une matrice d'association reliant chaque caractéristique de chaque modalité aux caractéristiques de la modalité initialement perturbée. Il s'agit d'un module d'interprétation construit pour comprendre les relations intermodales apprises par un modèle mVAE.

La stabilité est essentielle pour espérer garantir la reproductibilité des résultats. Après avoir examiné certaines origines de l'instabilité en apprentissage automatique, nous proposons des moyens d'atténuer ces sources de variabilité lors de l'interprétation des modèles mVAE à l'aide de la DAA. La variabilité épistémique (due à l'entraînement stochastique des réseaux de neurones) est gérée en répétant la procédure d'entraînement / interprétation du mVAE sur la même partition entraînement / test de la population, puis en assemblant leurs résultats. Il s'agit de la DAA régularisée (r-DAA). Afin de s'assurer que les associations observées ne sont pas spécifiques à la partition entraînement / test considérée par la r-DAA, nous proposons également d'utiliser une procédure de sélection par stabilité. Cela permet de conserver les associations sélectionnées de manière cohérente par la r-DAA sur plusieurs partitions de la population, ce qui permet de contrôler l'incertitude aléatorique.

Ces méthodes ont été présentées dans un contexte plus général, mais ont en fait été développées pour découvrir des associations cerveau-comportement reproductibles. Nous pensons que l'application de ces techniques à de grandes cohortes multimodales pourrait permettre de trouver des biomarqueurs pertinents. La DAA a donné lieu à une publication dans la conférence ISBI, présen-

tée en 2023 à Cartagena, en Colombie, et lors d'une session de posters au séminaire français IABM 2023 à Paris.
Dans la partie suivante, nous appliquons le module DAA à une cohorte transdiagnostique d'enfants à risque pour divers syndromes psychiatriques (dont l'autisme, l'hyperactivité et l'anxiété), présentant des symptômes partagés ou spécifiques de ces pathologies. Nous nous assurons d'obtenir des résultats d'associations reproductibles en appliquant les procédures de r-DAA et de sélection par stabilité. Nous discutons des paramètres expérimentaux associés et des résultats obtenus, ainsi que de quelques recommandations sur la façon de les interpréter, en particulier dans une perspective transdiagnostique.

Applications. Comme indiqué précédemment, le projet RDoC fournit des recommandations pour aborder correctement la complexité des maladies psychiatriques. Ces dernières années, la littérature transdiagnostique en psychopathologie s'est développée dans ce sens [125]. Certaines des ces approches transdiagnostiques récentes étudient un facteur psychopathologique général, appelé facteur p [53, 55]. Ce facteur p sous-tendrait des mécanismes communs à plusieurs syndromes psychiatriques ; il est généralement représenté comme une dimension unique qui agirait sur un ensemble de symptômes et, in fine, des troubles. L'objectif de ces études est de rechercher des corrélats neuronaux du facteur p parmi différents marqueurs biologiques, tels que l'imagerie ou la génétique.

La cohorte HBN [6] est une grande étude clinique multicentrique et transdisciplinaire. Elle comprend une grande diversité de type de mesures, y compris l'imagerie, et un ensemble complet d'évaluations psychologiques et cliniques pour mieux comprendre les troubles psychiatriques. Les critères d'inclusion ne dépendent pas du diagnostic, mais englobent plutôt une population à risque présentant des symptômes comportementaux notables. Plus précisément, les sujets ont été sélectionnés en fonction de la présence de structures comportementales liées au spectre autistique, à hyperactivité ou à d'autres troubles anxieux. Les diagnostics consensus ne sont pas disponibles pour la majorité des sujets inscrits dans la cohorte HBN. En tant que tel, l'ensemble de données permet d'étudier les différentes manifestations des syndromes psychiatriques au sein des données. Il permet notamment d'explorer les méthodes de découverte de nouveaux biomarqueurs et d'analyse dimensionnelle. À cet égard, HBN constitue un ensemble de données candidat très prometteur pour les études transdiagnostiques visant à identifier les biomarqueurs qui sous-tendent les processus biologiques dans l'ensemble des troubles. Nous proposons deux études de cette cohorte.

La première étude se concentrent principalement sur le spectre autistique. Des travaux précédents dans l'équipe ont proposé une stratification par un expert des sujets autistes annotés sur la base de différents scores de symptômes, et en s'appuyant principalement sur le score cognitif de déficience sociale SRS pour dissocier des groupes de contrôle de sous-groupes autistes. Nous avons réussi à reproduire ces étiquettes en utilisant une approche intégrative non supervisée en apprentissage profond avec un mVAE particulier appelé sMCVAE [14]. A aucun moment le réseau de neurones ne s'est vu proposé d'utiliser ces étiquettes pendant l'entraînement, mais pourtant il a été capable de reproduire une stratification similaire des sujets autistes dans l'espace latent spécifique

aux score cliniques. Le sMCVAE présente des avantages, notamment proposant une façon plus intégrée de faire de la stratification dans un espace principalement dirigé par de la variabilité des scores clinique mais comprenant un peu de variabilité d'imagerie. En revanche, le fait de ne pas modéliser un espace latent commun et d'imposer la même dimension latente pour tous les espaces latents spécifiques à chaque modalité ne constitue pas une bonne propriété pour interpréter le réseau de manière plus approfondie. Cette étude de réplication a été présentée pendant une session de posters virtuelle à OHBM 2021.

L'étude précédente se concentre principalement sur l'étude du spectre autistique. Dans cette deuxième étude, nous souhaitons découvrir des associations transdiagnostiques reproductibles entre les mesures de la surface corticale et les comportements symptomatiques impliqués dans différents troubles psychiatriques. Nous démontrons ici la capacité des outils méthodologiques précédemment présentés à découvrir de telles associations stables. Nous intégrons les résultats des questionnaires et les mesures corticales sur des régions d'intérêt en utilisant un mVAE, puis nous utilisons le cadre DAA pour interpréter ses représentations conjointes et ses associations en sortie. En particulier, nous montrons que cette procédure de r-DAA permet effectivement de gérer une part de l'instabilité des modèles dues à la variabilité épistémique et d'extraire des associations pertinentes. Encapsulé dans un cadre de sélection par stabilité, cet outil d'interprétation r-DAA garantit que les associations retenues sont généralisables au sein de notre population, et donc plus susceptibles d'être reproductibles. Nous montrons aussi que les modèles entrainés dans ce cadre sont munis de propriétés pertinentes en vue de leur interprétation : l'espace latent partagé contient de la variabilité d'intérêt, que lui permet de correctement modéliser ces structures de corrélation inter-vue. En revanche, la variabilité spécifique, incluant les effets confondants liée au site d'acquisition de l'image, sont contenus dans l'espace latent spécifique à l'imagerie. Celà démontre les capacités des mVAE à séparer les sources de variabilité lorsqu'ils modélisent à la fois un espace latent partagé entre les différentes modalités et un espace spécifique dédié à chaque modalité. En analysant les associations trouvées à travers différents scores de comportements corrélés à différents spectres psychiatriques, on identifie des régions de la surface corticale liées au facteur de psychopatholgie transdiagnostique : en l'occurence, les régions cingulaires (avec les paramètres d'aire et de courbure), ainsi que les cortex visuels primaires (avec le parametre d'épaisseur corticale). Ces contributions méthodologiques (r-DAA + stabilité par sélection) appliquées dans ce contexte ont été soumises à la revue Imaging Neuroscience et présentée durant le colloque IABM 2024 à Grenoble.

Bibliography

[1] Anissa Abi-Dargham, Scott J. Moeller, Farzana Ali, Christine DeLorenzo, Katharina Domschke, Guillermo Horga, Amandeep Jutla, Roman Kotov, Martin P. Paulus, Jose M. Rubio, Gerard Sanacora, Jeremy Veenstra-VanderWeele, and John H. Krystal. "Candidate biomarkers in psychiatric disorders: state of the field". en. In: *World Psychiatry* 22.2 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wps.21078, pp. 236–262. issn: 2051-5545. doi: 10.1002/wps.21078. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/wps.21078 (visited on 04/08/2024).

[2] Abubakar Abid, Martin J. Zhang, Vivek K. Bagaria, and James Zou. "Exploring patterns enriched in a dataset with contrastive principal component analysis". en. In: *Nature Communications* 9.1 (May 2018). Number: 1 Publisher: Nature Publishing Group, p. 2134. issn: 2041-1723. doi: 10.1038/s41467-018-04608-8. url: https://www.nature.com/articles/s41467-018-04608-8 (visited on 02/20/2024).

[3] Abubakar Abid and James Zou. *Contrastive Variational Autoencoder Enhances Salient Features*. arXiv:1902.04601 [cs, stat]. Feb. 2019. doi: 10.48550/arXiv.1902.04601. url: http://arxiv.org/abs/1902.04601 (visited on 02/20/2024).

[4] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. "Sanity Checks for Saliency Maps". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018. url: https://papers.nips.cc/paper_files/paper/2018/hash/294a8ed24b1ad22ec2e7efea049b8737-Abstract.html (visited on 03/08/2024).

[5] Aidas Aglinskas, Joshua K. Hartshorne, and Stefano Anzellotti. "Contrastive machine learning reveals the structure of neuroanatomical variation within autism". In: *Science* 376.6597 (June 2022). Publisher: American Association for the Advancement of Science, pp. 1070–1074. doi: 10.1126/science.abm2461. url: https://www.science.org/doi/10.1126/science.abm2461 (visited on 02/20/2024).

[6] Lindsay M. Alexander, Jasmine Escalera, et al. "An open resource for transdiagnostic research in pediatric mental health and learning disorders". en. In: *Scientific Data* 4.1 (Dec. 2017). Number: 1 Publisher: Nature Publishing Group, p. 170181. issn: 2052-4463. doi: 10.1038/sdata.2017.181. url: https://www.nature.com/articles/sdata2017181 (visited on 01/18/2024).

[7] Kate Allsopp, John Read, Rhiannon Corcoran, and Peter Kinderman. "Heterogeneity in psychiatric diagnostic classification". In: *Psychiatry Research* 279 (Sept. 2019), pp. 15–22. issn: 0165-1781. doi: 10.1016/j.psychres.2019.07.005. url: https://www.sciencedirect.com/science/article/pii/S0165178119309114 (visited on 03/12/2024).

[8] Laith Alzubaidi, Mohammed A. Fadhel, Omran Al-Shamma, Jinglan Zhang, J. Santamaría, Ye Duan, and Sameer R. Oleiwi. "Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study". en. In: *Applied Sciences* 10.13 (Jan. 2020). Number: 13 Publisher: Multidisciplinary Digital Publishing Institute, p. 4523. issn: 2076-3417. doi: 10.3390/app10134523. url: https://www.mdpi.com/2076-3417/10/13/4523 (visited on 01/18/2024).

[9] S.-I. Amari. "Learning Patterns and Pattern Sequences by Self-Organizing Nets of Threshold Elements". In: *IEEE Transactions on Computers* C-21.11 (Nov. 1972). Conference Name: IEEE Transactions on Computers, pp. 1197–1206. issn: 1557-9956. doi: 10.1109/T-C.1972.223477. url: https://ieeexplore.ieee.org/document/1672070 (visited on 03/04/2024).

[10] Shunichi Amari. "A Theory of Adaptive Pattern Classifiers". In: *IEEE Transactions on Electronic Computers* EC-16.3 (June 1967). Conference Name: IEEE Transactions on Electronic Computers, pp. 299–307. issn: 0367-7508. doi: 10.1109/PGEC.1967.264666. url: https://ieeexplore.ieee.org/document/4039068 (visited on 02/29/2024).

[11] Corentin Ambroise, Antoine Grigis, Edouard Duchesnay, and Vincent Frouin. "Multi-View Variational Autoencoders Allow for Interpretability Leveraging Digital Avatars: Application to the HBN Cohort". In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. ISSN: 1945-8452. Apr. 2023, pp. 1–5. doi: 10.1109/ISBI53787.2023.10230552. url: https://ieeexplore.ieee.org/document/10230552 (visited on 01/18/2024).

[12] Francesco Amico, Jonathan Stauber, Nikolaos Koutsouleris, and Thomas Frodl. "Anterior cingulate cortex gray matter abnormalities in adults with attention deficit hyperactivity disorder: A voxel-based morphometry study". In: *Psychiatry Research: Neuroimaging* 191.1 (Jan. 2011), pp. 31–35. issn: 0925-4927. doi: 10.1016/j.pscychresns.2010.08.011. url: https://www.sciencedirect.com/science/article/pii/S0925492710002933 (visited on 04/11/2024).

[13] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. "Deep Canonical Correlation Analysis". en. In: *Proceedings of the 30th International Conference on Machine Learning*. ISSN: 1938-7228. PMLR, May 2013, pp. 1247–1255. url: https://proceedings.mlr.press/v28/andrew13.html (visited on 03/19/2024).

[14] Luigi Antelmi, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. "Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data". en. In: *Proceedings of the 36th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, May 2019, pp. 302–311. url: https://proceedings.mlr.press/v97/antelmi19a.html (visited on 03/20/2024).

[15] Luigi Antelmi, Nicholas Ayache, Philippe Robert, Federica Ribaldi, Valentina Garibotto, Giovanni B. Frisoni, and Marco Lorenzi. *Combining Multi-Task Learning and Multi-Channel Variational Auto-Encoders to Exploit Datasets with Missing Observations -*

Application to Multi-Modal Neuroimaging Studies in Dementia. en. May 2021. url: https://inria.hal.science/hal-03114888 (visited on 04/17/2024).

[16] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. "VQA: Visual Question Answering". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 2425–2433. url: https://openaccess.thecvf.com/content_iccv_2015/html/Antol_VQA_Visual_Question_ICCV_2015_paper.html (visited on 03/26/2024).

[17] Martijn Arns, Hanneke van Dijk, Jurjen J. Luykx, Guido van Wingen, and Sebastian Olbrich. "Stratified psychiatry: Tomorrow's precision psychiatry?" In: *European Neuropsychopharmacology* 55 (Feb. 2022), pp. 14–19. issn: 0924-977X. doi: 10.1016/j.euroneuro.2021.10.863. url: https://www.sciencedirect.com/science/article/pii/S0924977X21016345 (visited on 03/12/2024).

[18] Devansh Arpit, Stanisław Jastrzębski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, and Simon Lacoste-Julien. "A Closer Look at Memorization in Deep Networks". en. In: *Proceedings of the 34th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2017, pp. 233–242. url: https://proceedings.mlr.press/v70/arpit17a.html (visited on 02/29/2024).

[19] James Atwood and Don Towsley. "Diffusion-Convolutional Neural Networks". In: *arXiv:1511.02136 [cs]* (July 2016). arXiv: 1511.02136. url: http://arxiv.org/abs/1511.02136 (visited on 12/02/2020).

[20] Pierre Auriau, Antoine Grigis, Benoit Dufumier, Robin Louiset, Joel Chavas, Pietro Gori, Jean-François Mangin, and Edouard Duchesnay. "Supervised diagnosis prediction from cortical sulci: toward the discovery of neurodevelopmental biomarkers in mental disorders". In: *21st IEEE International Symposium on Biomedical Imaging (ISBI 2024)*. Athènes, Greece: IEEE, May 2024. url: https://hal.science/hal-04494994 (visited on 04/17/2024).

[21] Francis R Bach and Michael I Jordan. *A Probabilistic Interpretation of Canonical Correlation Analysis*. en. Technical 688. University of California, Berkeley: Departement of Statistics, 2005.

[22] Monya Baker. "1,500 scientists lift the lid on reproducibility". en. In: *Nature* 533.7604 (May 2016). Publisher: Nature Publishing Group, pp. 452–454. issn: 1476-4687. doi: 10.1038/533452a. url: https://www.nature.com/articles/533452a (visited on 03/12/2024).

[23] Luca Baldassarre, Massimiliano Pontil, and Janaina Mourão-Miranda. "Sparsity Is Better with Stability: Combining Accuracy and Stability for Model Selection in Brain Decoding". English. In: *Frontiers in Neuroscience* 11 (Feb. 2017). Publisher: Frontiers. issn: 1662-453X. doi: 10.3389/fnins.2017.00062. url: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2017.00062/full (visited on 03/21/2024).

[24] Randall Balestriero, Mark Ibrahim, Vlad Sobal, Ari Morcos, Shashank Shekhar, Tom Goldstein, Florian Bordes, Adrien Bardes, Gregoire Mialon, Yuandong Tian, Avi Schwarzschild, Andrew Gordon Wilson, Jonas Geiping, Quentin Garrido, Pierre Fernandez, Amir Bar, Hamed Pirsiavash, Yann LeCun, and Micah Goldblum. *A Cookbook of Self-Supervised Learning*. arXiv:2304.12210 [cs]. June 2023. url: http://arxiv.org/abs/2304.12210 (visited on 09/13/2023).

[25] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. "Multimodal Machine Learning: A Survey and Taxonomy". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2 (Feb. 2019). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 423–443. issn: 1939-3539. doi: 10.1109/TPAMI.2018.2798607. url: https://ieeexplore.ieee.org/document/8269806 (visited on 03/19/2024).

[26] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. "BEiT: BERT Pre-Training of Image Transformers". en. In: *The Tenth International Conference on Learning Representations, {ICLR} 2022, Virtual Event, April 25-29, 2022*. Oct. 2021. url: https://openreview.net/forum?id=p-BhZSz59o4 (visited on 03/08/2024).

[27] Hugo Barbaroux, Xinyang Feng, Jie Yang, Andrew F. Laine, and Elsa D. Angelini. "Encoding Human Cortex Using Spherical CNNs - A Study on Alzheimer's Disease Classification". en. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Iowa City, IA, USA: IEEE, Apr. 2020, pp. 1322–1325. isbn: 978-1-5386-9330-8. doi: 10.1109/ISBI45749.2020.9098353. url: https://ieeexplore.ieee.org/document/9098353/ (visited on 01/18/2024).

[28] Adrien Bardes, Jean Ponce, and Yann LeCun. "VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning". en. In: *The Tenth International Conference on Learning Representations, {ICLR} 2022, Virtual Event, April 25-29, 2022*. Oct. 2021. url: https://openreview.net/forum?id=xm6YD62D1Ub (visited on 03/07/2024).

[29] Arnab Barua, Mobyen Uddin Ahmed, and Shahina Begum. "A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions". In: *IEEE Access* 11 (2023). Conference Name: IEEE Access, pp. 14804–14831. issn: 2169-3536. doi: 10.1109/ACCESS.2023.3243854. url: https://ieeexplore.ieee.org/document/10041115 (visited on 03/19/2024).

[30] Vishnu M. Bashyam, Jimit Doshi, et al. "Deep Generative Medical Image Harmonization for Improving Cross-Site Generalization in Deep Learning Predictors". en. In: *Journal of Magnetic Resonance Imaging* 55.3 (2022). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.27908, pp. 908–916. issn: 1522-2586. doi: 10.1002/jmri.27908. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.27908 (visited on 02/19/2024).

[31] Frida Bayard, Charlotte Nymberg Thunell, et al. “Distinct brain structure and behavior related to ADHD and conduct disorder traits”. en. In: *Molecular Psychiatry* 25.11 (Nov. 2020). Publisher: Nature Publishing Group, pp. 3020–3033. issn: 1476-5578. doi: 10.1038/s41380-018-0202-6. url: https://www.nature.com/articles/s41380-018-0202-6 (visited on 04/11/2024).

[32] H. Benali and Y. Cointepas. “BrainVISA: software platform for visualization and analysis of multi-modality brain data”. In: *Academia Press* (Jan. 2001). url: https://www.academia.edu/12445758/BrainVISA_software_platform_for_visualization_and_analysis_of_multi_modality_brain_data (visited on 03/28/2024).

[33] Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. *Graph Neural Networks in Network Neuroscience*. arXiv:2106.03535 [cs, q-bio]. Sept. 2022. doi: 10.48550/arXiv.2106.03535. url: http://arxiv.org/abs/2106.03535 (visited on 04/03/2024).

[34] Erin D. Bigler. “Structural Image Analysis of the Brain in Neuropsychology Using Magnetic Resonance Imaging (MRI) Techniques”. en. In: *Neuropsychology Review* 25.3 (Sept. 2015), pp. 224–249. issn: 1573-6660. doi: 10.1007/s11065-015-9290-0. url: https://doi.org/10.1007/s11065-015-9290-0 (visited on 04/07/2024).

[35] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. en. Springer, 2006. url: https://link.springer.com/book/9780387310732 (visited on 03/07/2024).

[36] Rishi Bommasani, Drew A. Hudson, et al. *On the Opportunities and Risks of Foundation Models*. arXiv:2108.07258 [cs]. July 2022. doi: 10.48550/arXiv.2108.07258. url: http://arxiv.org/abs/2108.07258 (visited on 01/26/2024).

[37] Simona Bottani, Ninon Burgos, Aurélien Maire, Dario Saracino, Sebastian Ströer, Didier Dormont, and Olivier Colliot. “Evaluation of MRI-based machine learning approaches for computer-aided diagnosis of dementia in a clinical data warehouse”. In: *Medical Image Analysis* 89 (Oct. 2023), p. 102903. issn: 1361-8415. doi: 10.1016/j.media.2023.102903. url: https://www.sciencedirect.com/science/article/pii/S1361841523001639 (visited on 04/09/2024).

[38] Nicolas Boulant, Lionel Quettier, et al. “Commissioning of the Iseult CEA 11.7 T whole-body MRI: current status, gradient–magnet interaction tests and first imaging experience”. en. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 36.2 (Apr. 2023), pp. 175–189. issn: 1352-8661. doi: 10.1007/s10334-023-01063-5. url: https://doi.org/10.1007/s10334-023-01063-5 (visited on 04/03/2024).

[39] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. “Signature Verification using a "Siamese" Time Delay Neural Network”. In: *Advances in Neural Information Processing Systems*. Vol. 6. Morgan-Kaufmann, 1993. url: https://proceedings.neurips.cc/paper/1993/hash/288cc0ff022877bd3df94bc9360b9c5d-Abstract.html (visited on 03/07/2024).

[40] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (July 2017). Conference Name: IEEE Signal Processing Magazine, pp. 18–42. issn: 1558-0792. doi: 10.1109/MSP.2017.2693418. url: https://ieeexplore.ieee.org/document/7974879 (visited on 02/19/2024).

[41] Tom Brown, Benjamin Mann, et al. "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901. url: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html (visited on 04/02/2024).

[42] Joan Bruna, Christian Szegedy, Ilya Sutskever, Ian Goodfellow, Wojciech Zaremba, Rob Fergus, and Dumitru Erhan. "Intriguing properties of neural networks". en. In: *2nd International Conference on Learning Representations, {ICLR} 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. Dec. 2013. url: https://openreview.net/forum?id=kklr_MTHMRQjG (visited on 03/05/2024).

[43] STEPHEN G. BRUSH. "History of the Lenz-Ising Model". In: *Reviews of Modern Physics* 39.4 (Oct. 1967). Publisher: American Physical Society, pp. 883–893. doi: 10.1103/RevModPhys.39.883. url: https://link.aps.org/doi/10.1103/RevModPhys.39.883 (visited on 03/04/2024).

[44] Anthony S. Bryk and Stephen W. Raudenbush. *Hierarchical linear models: Applications and data analysis methods*. Hierarchical linear models: Applications and data analysis methods. Pages: xvi, 265. Thousand Oaks, CA, US: Sage Publications, Inc, 1992. isbn: 978-0-8039-4627-9.

[45] Christopher P. Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. "Understanding disentangling in β-VAE". In: *arXiv:1804.03599 [cs, stat]* (Apr. 2018). arXiv: 1804.03599. url: http://arxiv.org/abs/1804.03599 (visited on 04/09/2021).

[46] George Bush, Jean A Frazier, Scott L Rauch, Larry J Seidman, Paul J Whalen, Michael A Jenike, Bruce R Rosen, and Joseph Biederman. "Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the counting stroop". In: *Biological Psychiatry* 45.12 (June 1999), pp. 1542–1552. issn: 0006-3223. doi: 10.1016/S0006-3223(99)00083-9. url: https://www.sciencedirect.com/science/article/pii/S0006322399000839 (visited on 02/12/2024).

[47] George Bush, Phan Luu, and Michael I. Posner. "Cognitive and emotional influences in anterior cingulate cortex". In: *Trends in Cognitive Sciences* 4.6 (June 2000), pp. 215–222. issn: 1364-6613. doi: 10.1016/S1364-6613(00)01483-2. url: https://www.sciencedirect.com/science/article/pii/S1364661300014832 (visited on 02/09/2024).

[48] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, Adrian Cortes, Samantha Welsh, Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi Allen, Peter Donnelly, and Jonathan Marchini. "The UK Biobank resource with deep phenotyping and genomic data". en. In: *Nature* 562.7726 (Oct. 2018). Publisher: Nature Publishing Group, pp. 203–209. issn: 1476-4687. doi: 10.1038/s41586-018-0579-z. url: https://www.nature.com/articles/s41586-018-0579-z (visited on 04/08/2024).

[49] Mathilde Caron, Hugo Touvron, Ishan Misra, Herve Jegou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. "Emerging Properties in Self-Supervised Vision Transformers". en. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9630–9640. isbn: 978-1-66542-812-5. doi: 10.1109/ICCV48922.2021.00951. url: https://ieeexplore.ieee.org/document/9709990/ (visited on 01/18/2024).

[50] J. Douglas Carroll. "Generalization of Canonical Correlation Analysis to Three of More Sets of Variables". In: *American Psychological Association* (1968). doi: 10.1037/e473742008-115.

[51] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. "Machine Learning Interpretability: A Survey on Methods and Metrics". en. In: *Electronics* 8.8 (Aug. 2019). Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, p. 832. issn: 2079-9292. doi: 10.3390/electronics8080832. url: https://www.mdpi.com/2079-9292/8/8/832 (visited on 03/05/2024).

[52] B. J. Casey, Tariq Cannonier, et al. "The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites". In: *Developmental Cognitive Neuroscience*. The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy 32 (Aug. 2018), pp. 43–54. issn: 1878-9293. doi: 10.1016/j.dcn.2018.03.001. url: https://www.sciencedirect.com/science/article/pii/S1878929317301214 (visited on 03/13/2024).

[53] Avshalom Caspi, Renate M. Houts, Daniel W. Belsky, Sidra J. Goldman-Mellor, HonaLee Harrington, Salomon Israel, Madeline H. Meier, Sandhya Ramrakha, Idan Shalev, Richie Poulton, and Terrie E. Moffitt. "The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders?" en. In: *Clinical Psychological Science* 2.2 (Mar. 2014). Publisher: SAGE Publications Inc, pp. 119–137. issn: 2167-7026. doi: 10.1177/2167702613497473. url: https://doi.org/10.1177/2167702613497473 (visited on 01/31/2024).

[54] Avshalom Caspi, Renate M. Houts, Helen L. Fisher, Andrea Danese, and Terrie E. Moffitt. "The General Factor of Psychopathology (p): Choosing Among Competing Models and Interpreting p". en. In: *Clinical Psychological Science* 12.1 (Jan. 2024). Publisher: SAGE Publications Inc, pp. 53–82. issn: 2167-7026. doi: 10.1177/21677026221147872. url: https://doi.org/10.1177/21677026221147872 (visited on 04/11/2024).

[55] Avshalom Caspi and Terrie E. Moffitt. "All for One and One for All: Mental Disorders in One Dimension". In: *American Journal of Psychiatry* 175.9 (Sept. 2018). Publisher: American Psychiatric Publishing, pp. 831–844. issn: 0002-953X. doi: 10.1176/appi.ajp.2018.17121383. url: https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2018.17121383 (visited on 03/13/2024).

[56] Simon Caton and Christian Haas. *Fairness in Machine Learning: A Survey*. arXiv:2010.04053 [cs, stat]. Oct. 2020. doi: 10.48550/arXiv.2010.04053. url: http://arxiv.org/abs/2010.04053 (visited on 03/05/2024).

[57] Hamza Chegraoui, Vincent Guillemot, Amine Rebei, Arnaud Gloaguen, Jacques Grill, Cathy Philippe, and Vincent Frouin. "Integrating multiomics and prior knowledge: a study of the Graphnet penalty impact". In: *Bioinformatics* 39.8 (Aug. 2023), btad454. issn: 1367-4811. doi: 10.1093/bioinformatics/btad454. url: https://doi.org/10.1093/bioinformatics/btad454 (visited on 10/02/2023).

[58] Kumar Chellapilla, Sidd Puri, and Patrice Simard. "High Performance Convolutional Neural Networks for Document Processing". en. In: *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, Oct. 2006. url: https://inria.hal.science/inria-00112631 (visited on 02/29/2024).

[59] Ricky T. Q. Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. "Isolating Sources of Disentanglement in Variational Autoencoders". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018. url: https://proceedings.neurips.cc/paper_files/paper/2018/hash/1ee3dfcd8a0645a25a35977997223d22-Abstract.html (visited on 03/15/2024).

[60] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. "A Simple Framework for Contrastive Learning of Visual Representations". en. In: *Proceedings of the 37th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, Nov. 2020, pp. 1597–1607. url: https://proceedings.mlr.press/v119/chen20j.html (visited on 01/18/2024).

[61] Jiale Cheng, Xin Zhang, Hao Ni, Chenyang Li, Xiangmin Xu, Zhengwang Wu, Li Wang, Weili Lin, and Gang Li. "Path Signature Neural Network of Cortical Features for Prediction of Infant Cognitive Scores". In: *IEEE Transactions on Medical Imaging* 41.7 (July 2022). Conference Name: IEEE Transactions on Medical Imaging, pp. 1665–1676. issn: 1558-254X. doi: 10.1109/TMI.2022.3147690. url: https://ieeexplore.ieee.org/document/9696321 (visited on 04/03/2024).

[62] Jieyu Cheng, Adrian V. Dalca, Bruce Fischl, and Lilla Zöllei. "Cortical surface registration using unsupervised learning". In: *NeuroImage* 221 (Nov. 2020), p. 117161. issn: 1053-8119. doi: 10.1016/j.neuroimage.2020.117161. url: https://www.sciencedirect.com/science/article/pii/S1053811920306479 (visited on 04/03/2024).

[63] Yi-Ling Chien, Yu-Chieh Chen, and Susan Shur-Fen Gau. "Altered cingulate structures and the associations with social awareness deficits and CNTNAP2 gene in autism spectrum disorder". In: *NeuroImage: Clinical* 31 (Jan. 2021), p. 102729. issn: 2213-1582. doi: 10.1016/j.nicl.2021.102729. url: https://www.sciencedirect.com/science/article/pii/S221315822100173X (visited on 04/11/2024).

[64] Fang-Ying Chiu and Yun Yen. "Imaging biomarkers for clinical applications in neuro-oncology: current status and future perspectives". In: *Biomarker Research* 11.1 (Mar. 2023), p. 35. issn: 2050-7771. doi: 10.1186/s40364-023-00476-7. url: https://doi.org/10.1186/s40364-023-00476-7 (visited on 04/05/2024).

[65] S. Chopra, R. Hadsell, and Y. LeCun. "Learning a similarity metric discriminatively, with application to face verification". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. ISSN: 1063-6919. June 2005, 539–546 vol. 1. doi: 10.1109/CVPR.2005.202. url: https://ieeexplore.ieee.org/document/1467314 (visited on 03/07/2024).

[66] Anwesa Choudhuri, Ashok Vardhan Makkuva, Ranvir Rana, Sewoong Oh, Girish Chowdhary, and Alexander Schwing. *Towards Principled Objectives for Contrastive Disentanglement*. en. Sept. 2019. url: https://openreview.net/forum?id=B1lPETVFPS (visited on 02/20/2024).

[67] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. arXiv:1412.3555 [cs]. Dec. 2014. doi: 10.48550/arXiv.1412.3555. url: http://arxiv.org/abs/1412.3555 (visited on 03/04/2024).

[68] Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca M Gambardella, and Jurgen Schmidhuber. "Flexible, High Performance Convolutional Neural Networks for Image Classification". en. In: *{IJCAI} 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*. 2011.

[69] Brett A. Clementz, John A. Sweeney, Jordan P. Hamm, Elena I. Ivleva, Lauren E. Ethridge, Godfrey D. Pearlson, Matcheri S. Keshavan, and Carol A. Tamminga. "Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers". In: *American Journal of Psychiatry* 173.4 (Apr. 2016). Publisher: American Psychiatric Publishing, pp. 373–384. issn: 0002-953X. doi: 10.1176/appi.ajp.2015.14091200. url: https://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.2015.14091200 (visited on 04/11/2024).

[70] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. "Gauge Equivariant Convolutional Networks and the Icosahedral CNN". en. In: *Proceedings of the 36th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, May 2019, pp. 1321–1330. url: https://proceedings.mlr.press/v97/cohen19d.html (visited on 01/18/2024).

[71] James H. Cole and Katja Franke. "Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers". In: *Trends in Neurosciences* 40.12 (Dec. 2017), pp. 681–690. issn: 0166-2236. doi: 10.1016/j.tins.2017.10.001. url: https://www.sciencedirect.com/science/article/pii/S016622361730187X (visited on 04/03/2024).

[72] ENIGMA Consortium. "ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries". In: *Translational Psychiatry* (Mar. 2020). doi: 10.1038/s41398-020-0705-1.

[73] ENIGMA Consortium. "The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data". In: *Brain Imaging and Behavior* (June 2014). doi: 10.1007/s11682-013-9269-5.

[74] John N. Constantino, Sandra A. Davis, Richard D. Todd, Matthew K. Schindler, Maggie M. Gross, Susan L. Brophy, Lisa M. Metzger, Christiana S. Shoushtari, Reagan Splinter, and Wendy Reich. "Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised". en. In: *Journal of Autism and Developmental Disorders* 33.4 (Aug. 2003), pp. 427–433. issn: 1573-3432. doi: 10.1023/A:1025014929212. url: https://doi.org/10.1023/A:1025014929212 (visited on 04/09/2024).

[75] Moheb Costandi. *Neuroplasticity*. eng. MIT Press essential knowledge series. OCLC: 987683015. Cambridge, Massachusetts: The MIT Press, 2016. isbn: 978-0-262-52933-4.

[76] D. R. Cox. "Regression Models and Life-Tables". en. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (1972). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1972.tb00899.x, pp. 187–202. issn: 2517-6161. doi: 10.1111/j.2517-6161.1972.tb00899.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1972.tb00899.x (visited on 03/26/2024).

[77] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press, 2000. isbn: 978-0-521-78019-3. doi: 10.1017/CBO9780511801389. url: https://www.cambridge.org/core/books/an-introduction-to-support-vector-machines-and-other-kernelbased-learning-methods/A6A6F4084056A4B23F88648DDBFDD6FC (visited on 03/14/2024).

[78] Marc-Antoine Crocq. "A history of anxiety: from Hippocrates to DSM". In: *Dialogues in Clinical Neuroscience* 17.3 (Sept. 2015). Publisher: Taylor & Francis _eprint: https://doi.org/10.31887/DCNS.2015.17.3/macrocq, pp. 319–325. issn: null. doi: 10.31887/DCNS.2015.17.3/macrocq. url: https://doi.org/10.31887/DCNS.2015.17.3/macrocq (visited on 03/12/2024).

[79] Rodrigo Santa Cruz, Leo Lebrat, Pierrick Bourgeat, Clinton Fookes, Jurgen Fripp, and Olivier Salvado. "DeepCSR: A 3D Deep Learning Approach for Cortical Surface Reconstruction". In: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. ISSN: 2642-9381. Jan. 2021, pp. 806–815. doi: 10.1109/WACV48630.2021.00085. url: https://ieeexplore.ieee.org/document/9423272 (visited on 04/03/2024).

[80] Guillem Cucurull, Konrad Wagstyl, Arantxa Casanova, Petar Veličković, Estrid Jakobsen, Michal Drozdzal, Adriana Romero, Alan Evans, and Yoshua Bengio. "Convolutional neural networks for mesh-based parcellation of the cerebral cortex". en. In: *1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands*. 2018. url: https://openreview.net/forum?id=rkKvBAiiz (visited on 04/03/2024).

[81] Jeffrey Cummings and Jefferson Kinney. "Biomarkers for Alzheimer's Disease: Context of Use, Qualification, and Roadmap for Clinical Implementation". en. In: *Medicina* 58.7 (July 2022). Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, p. 952. issn: 1648-9144. doi: 10.3390/medicina58070952. url: https://www.mdpi.com/1648-9144/58/7/952 (visited on 04/05/2024).

[82] G. Cybenko. "Approximation by superpositions of a sigmoidal function". en. In: *Mathematics of Control, Signals and Systems* 2.4 (Dec. 1989), pp. 303–314. issn: 1435-568X. doi: 10.1007/BF02551274. url: https://doi.org/10.1007/BF02551274 (visited on 02/29/2024).

[83] A.-M. D'Cruz, M. W. Mosconi, M. E. Ragozzino, E. H. Cook, and J. A. Sweeney. "Alterations in the functional neural circuitry supporting flexible choice behavior in autism spectrum disorders". en. In: *Translational Psychiatry* 6.10 (Oct. 2016). Publisher: Nature Publishing Group, e916–e916. issn: 2158-3188. doi: 10.1038/tp.2016.161. url: https://www.nature.com/articles/tp2016161 (visited on 04/11/2024).

[84] Kamalaker Dadi, Gaël Varoquaux, Josselin Houenou, Danilo Bzdok, Bertrand Thirion, and Denis Engemann. "Population modeling with machine learning can enhance measures of mental health". In: *GigaScience* 10.10 (Oct. 2021), giab071. issn: 2047-217X. doi: 10.1093/gigascience/giab071. url: https://doi.org/10.1093/gigascience/giab071 (visited on 01/26/2024).

[85] Simon Dahan, Logan Z. J. Williams, Daniel Rueckert, and Emma C. Robinson. "Improving Phenotype Prediction Using Long-Range Spatio-Temporal Dynamics of Functional Connectivity". en. In: *Machine Learning in Clinical Neuroimaging*. Ed. by Ahmed Abdulkadir, Seyed Mostafa Kia, Mohamad Habes, Vinod Kumar, Jane Maryam Rondina, Chantal Tax, and Thomas Wolfers. Cham: Springer International Publishing, 2021, pp. 145–154. isbn: 978-3-030-87586-2. doi: 10.1007/978-3-030-87586-2_15.

[86] Anders M. Dale, Bruce Fischl, and Martin I. Sereno. "Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction". In: *NeuroImage* 9.2 (Feb. 1999), pp. 179–194. issn: 1053-8119. doi: 10.1006/nimg.1998.0395. url: https://www.sciencedirect.com/science/article/pii/S1053811998903950 (visited on 01/18/2024).

[87] Tim Dalgleish, Melissa Black, David Johnston, and Anna Bevan. "Transdiagnostic approaches to mental health problems: Current status and future directions." en. In: *Journal of Consulting and Clinical Psychology* 88.3 (Mar. 2020), pp. 179–195. issn: 1939-2117, 0022-006X. doi: 10.1037/ccp0000482. url: https://doi.apa.org/doi/10.1037/ccp0000482 (visited on 04/11/2024).

[88] Imant Daunhawer, Thomas M. Sutter, Kieran Chin-Cheong, Emanuele Palumbo, and Julia E. Vogt. "On the Limitations of Multimodal VAEs". en. In: *The Tenth International Conference on Learning Representations, {ICLR} 2022, Virtual Event, April 25-29, 2022*. Oct. 2021. url: https://openreview.net/forum?id=w-CPUXXrAj (visited on 03/19/2024).

[89] Imant Daunhawer, Thomas M. Sutter, Ricards Marcinkevics, and Julia E. Vogt. "Self-supervised Disentanglement of Modality-Specific and Shared Factors Improves Multimodal Generative Models". In: *Pattern Recognition: 42nd DAGM German Conference, DAGM GCPR 2020, Tübingen, Germany, September 28 – October 1, 2020, Proceedings*. Berlin, Heidelberg: Springer-Verlag, Sept. 2020, pp. 459–473. isbn: 978-3-030-71277-8. doi: 10.1007/978-3-030-71278-5_33. url: https://doi.org/10.1007/978-3-030-71278-5_33 (visited on 03/19/2024).

[90] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering". In: *arXiv:1606.09375 [cs, stat]* (Feb. 2017). arXiv: 1606.09375. url: http://arxiv.org/abs/1606.09375 (visited on 12/15/2020).

[91] Michaël Defferrard, Martino Milani, Frédérick Gusset, and Nathanaël Perraudin. "DeepSphere: a graph-based spherical CNN". en. In: *8th International Conference on Learning Representations, {ICLR} 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. Sept. 2019. url: https://openreview.net/forum?id=B1e3OlStPB (visited on 01/18/2024).

[92] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. ISSN: 1063-6919. June 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848. url: https://ieeexplore.ieee.org/document/5206848 (visited on 02/29/2024).

[93] Thomas F. Denson, William C. Pedersen, Jaclyn Ronquillo, and Anirvan S. Nandy. "The Angry Brain: Neural Correlates of Anger, Angry Rumination, and Aggressive Personality". In: *Journal of Cognitive Neuroscience* 21.4 (Apr. 2009), pp. 734–744. issn: 0898-929X. doi: 10.1162/jocn.2009.21051. url: https://doi.org/10.1162/jocn.2009.21051 (visited on 02/09/2024).

[94] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". In: *NeuroImage* 31.3 (July 2006), pp. 968–980. issn: 1053-8119. doi: 10.1016/j.

neuroimage.2006.01.021. url: https://www.sciencedirect.com/science/article/pii/S1053811906000437 (visited on 03/31/2024).

[95] Nandita M. deSouza, Eric Achten, Angel Alberich-Bayarri, Fabian Bamberg, Ronald Boellaard, Olivier Clément, Laure Fournier, Ferdia Gallagher, Xavier Golay, Claus Peter Heussel, Edward F. Jackson, Rashindra Manniesing, Marius E. Mayerhofer, Emanuele Neri, James O'Connor, Kader Karli Oguz, Anders Persson, Marion Smits, Edwin J. R. van Beek, Christoph J. Zech, and European Society of Radiology. "Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL* subcommittee of the European Society of Radiology (ESR)". In: *Insights into Imaging* 10.1 (Aug. 2019), p. 87. issn: 1869-4101. doi: 10.1186/s13244-019-0764-0. url: https://doi.org/10.1186/s13244-019-0764-0 (visited on 04/04/2024).

[96] Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature". In: *NeuroImage* 53.1 (Oct. 2010), pp. 1–15. issn: 1053-8119. doi: 10.1016/j.neuroimage.2010.06.010. url: https://www.sciencedirect.com/science/article/pii/S1053811910008542 (visited on 03/31/2024).

[97] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv:1810.04805 [cs]. May 2019. doi: 10.48550/arXiv.1810.04805. url: http://arxiv.org/abs/1810.04805 (visited on 03/06/2024).

[98] A. Di Martino, C.-G. Yan, et al. "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism". en. In: *Molecular Psychiatry* 19.6 (June 2014). Publisher: Nature Publishing Group, pp. 659–667. issn: 1476-5578. doi: 10.1038/mp.2013.78. url: https://www.nature.com/articles/mp201378 (visited on 03/13/2024).

[99] Adriana Di Martino, David O'Connor, et al. "Enhancing studies of the connectome in autism using the autism brain imaging data exchange II". en. In: *Scientific Data* 4.1 (Mar. 2017). Publisher: Nature Publishing Group, p. 170010. issn: 2052-4463. doi: 10.1038/sdata.2017.10. url: https://www.nature.com/articles/sdata201710 (visited on 03/13/2024).

[100] Nicola K. Dinsdale, Mark Jenkinson, and Ana I. L. Namburete. "Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal". In: *NeuroImage* 228 (Mar. 2021), p. 117689. issn: 1053-8119. doi: 10.1016/j.neuroimage.2020.117689. url: https://www.sciencedirect.com/science/article/pii/S1053811920311745 (visited on 02/19/2024).

[101] Finale Doshi-Velez and Been Kim. *Towards A Rigorous Science of Interpretable Machine Learning*. arXiv:1702.08608 [cs, stat]. Mar. 2017. doi: 10.48550/arXiv.1702.08608. url: http://arxiv.org/abs/1702.08608 (visited on 03/05/2024).

[102] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". en. In: *9th International Conference on Learning Representations, {ICLR} 2021, Virtual Event, Austria, May 3-7, 2021*. Oct. 2020. url: https://openreview.net/forum?id=YicbFdNTTy (visited on 03/04/2024).

[103] Benoit Dufumier, Pietro Gori, Sara Petiton, Robin Louiset, Jean-François Mangin, Antoine Grigis, and Edouard Duchesnay. "Exploring the potential of representation and transfer learning for anatomical neuroimaging: application to psychiatry". In: *NeuroImage* (Feb. 2024). url: https://hal.science/hal-04436585 (visited on 03/28/2024).

[104] Benoit Dufumier, Pietro Gori, Julie Victor, Antoine Grigis, Michele Wessa, Paolo Brambilla, Pauline Favre, Mircea Polosan, Colm McDonald, Camille Marie Piguet, Mary Phillips, Lisa Eyler, and Edouard Duchesnay. "Contrastive Learning with Continuous Proxy Meta-data for 3D MRI Classification". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Ed. by Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 58–68. isbn: 978-3-030-87196-3. doi: 10.1007/978-3-030-87196-3_6.

[105] Benoit Dufumier, Antoine Grigis, Julie Victor, Corentin Ambroise, Vincent Frouin, and Edouard Duchesnay. "OpenBHB: a Large-Scale Multi-Site Brain MRI Data-set for Age Prediction and Debiasing". In: *NeuroImage* 263 (Nov. 2022), p. 119637. issn: 1053-8119. doi: 10.1016/j.neuroimage.2022.119637. url: https://www.sciencedirect.com/science/article/pii/S1053811922007522 (visited on 01/18/2024).

[106] Christine Ecker, Charlotte M. Pretzsch, et al. "Interindividual Differences in Cortical Thickness and Their Genomic Underpinnings in Autism Spectrum Disorder". In: *American Journal of Psychiatry* 179.3 (Mar. 2022). Publisher: American Psychiatric Publishing, pp. 242–254. issn: 0002-953X. doi: 10.1176/appi.ajp.2021.20050630. url: https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2021.20050630 (visited on 02/12/2024).

[107] Fabian Eitel, Marc-André Schulz, Moritz Seiler, Henrik Walter, and Kerstin Ritter. "Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research". In: *Experimental Neurology* 339 (May 2021), p. 113608. issn: 0014-4886. doi: 10.1016/j.expneurol.2021.113608. url: https://www.sciencedirect.com/science/article/pii/S0014488621000133 (visited on 01/26/2024).

[108] Maxwell L. Elliott, Adrienne Romer, Annchen R. Knodt, and Ahmad R. Hariri. "A Connectome-wide Functional Signature of Transdiagnostic Risk for Mental Illness". In: *Biological Psychiatry*. Translating Biology to Treatment in Schizophrenia 84.6 (Sept. 2018), pp. 452–459. issn: 0006-3223. doi: 10.1016/j.biopsych.2018.03.012. url: https://www.sciencedirect.com/science/article/pii/S0006322318314161 (visited on 02/01/2024).

[109] Kristian M. Eschenburg, Thomas J. Grabowski, and David R. Haynor. "Learning Cortical Parcellations Using Graph Neural Networks". English. In: *Frontiers in Neuroscience* 15 (Dec. 2021). Publisher: Frontiers. issn: 1662-453X. doi: 10.3389/fnins.2021.797500. url: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2021.797500/full (visited on 04/03/2024).

[110] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. "Learning SO(3) Equivariant Representations with Spherical CNNs". In: *Computer Vision - {ECCV} 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part {XIII}*. 2018, pp. 52–68. url: https://openaccess.thecvf.com/content_ECCV_2018/html/Carlos_Esteves_Learning_SO3_Equivariant_ECCV_2018_paper.html (visited on 01/26/2024).

[111] Damien A. Fair, Deepti Bathula, Molly A. Nikolas, and Joel T. Nigg. "Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD". In: *Proceedings of the National Academy of Sciences* 109.17 (Apr. 2012). Publisher: Proceedings of the National Academy of Sciences, pp. 6769–6774. doi: 10.1073/pnas.1115365109. url: https://www.pnas.org/doi/full/10.1073/pnas.1115365109 (visited on 03/12/2024).

[112] Peter Falkai, Andrea Schmitt, and Nancy Andreasen. "Forty years of structural brain imaging in mental disorders: is it clinically useful or not?" In: *Dialogues in Clinical Neuroscience* 20.3 (Sept. 2018). Publisher: Taylor & Francis _eprint: https://doi.org/10.31887/DCNS.2018.20.3/pfalkai, pp. 179–186. issn: null. doi: 10.31887/DCNS.2018.20.3/pfalkai. url: https://doi.org/10.31887/DCNS.2018.20.3/pfalkai (visited on 04/05/2024).

[113] Abdulah Fawaz, Logan Z. J. Williams, Amir Alansary, Cher Bass, Karthik Gopinath, Mariana Da Silva, Simon Dahan, Chris Adamson, Bonnie Alexander, Deanne Thompson, Gareth Ball, Christian Desrosiers, Hervé Lombaert, Daniel Rueckert, A. David Edwards, and Emma C. Robinson. *Benchmarking Geometric Deep Learning for Cortical Segmentation and Neurodevelopmental Phenotype Prediction*. en. Dec. 2021. doi: 10.1101/2021.12.01.470730. url: http://biorxiv.org/lookup/doi/10.1101/2021.12.01.470730 (visited on 03/28/2024).

[114] Eric Feczko, Oscar Miranda-Dominguez, Mollie Marr, Alice M. Graham, Joel T. Nigg, and Damien A. Fair. "The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes". In: *Trends in Cognitive Sciences* 23.7 (July 2019), pp. 584–601. issn: 1364-6613. doi: 10.1016/j.tics.2019.03.009. url: https://www.sciencedirect.com/science/article/pii/S1364661319300920 (visited on 03/12/2024).

[115] K. Feldker, C. Y. Heitmann, P. Neumeister, S. V. Tupak, E. Schrammen, R. Moeck, P. Zwitserlood, M. Bruchmann, and T. Straube. "Transdiagnostic brain responses to disorder-related threat across four psychiatric disorders". en. In: *Psychological Medicine* 47.4 (Mar. 2017). Publisher: Cambridge University Press, pp. 730–743. issn: 0033-2917, 1469-8978. doi: 10.1017/S0033291716002634. url: https://

www . cambridge . org / core / journals / psychological - medicine / article / transdiagnostic - brain - responses - to - disorderrelated - threat - across - four - psychiatric - disorders / B53A02647B8258019725E4E569086B83 (visited on 02/01/2024).

[116] Xinyang Feng, Jie Yang, Andrew F. Laine, and Elsa D. Angelini. *Discriminative analysis of the human cortex using spherical CNNs - a study on Alzheimer's disease diagnosis*. arXiv:1812.07749 [cs]. Dec. 2018. doi: 10 . 48550 / arXiv . 1812 . 07749. url: http : //arxiv.org/abs/1812.07749 (visited on 04/03/2024).

[117] Santiago Fernández, Alex Graves, and Jürgen Schmidhuber. "An application of recurrent neural networks to discriminative keyword spotting". In: *Proceedings of the 17th international conference on Artificial neural networks*. ICANN'07. Berlin, Heidelberg: Springer-Verlag, Sept. 2007, pp. 220–229. isbn: 978-3-540-74693-5. (Visited on 03/04/2024).

[118] Virginia Fernández, Cristina Llinares-Benadero, and Víctor Borrell. "Cerebral cortex expansion and folding: what have we learned?" In: *The EMBO Journal* 35.10 (May 2016). Num Pages: 1044 Publisher: John Wiley & Sons, Ltd, pp. 1021–1044. issn: 0261-4189. doi: 10.15252/embj.201593701. url: https://www.embopress.org/doi/full/10.15252/embj.201593701 (visited on 04/17/2024).

[119] Clara Fischer, Gregory Operto, S. Laguitton, Matthieu Perrot, I. Denghien, Denis Rivière, and Jean-François Mangin. "Morphologist 2012: the new morphological pipeline of BrainVISA". In: *Proceedings of the 18th Annual Meeting of the Organization for Human Brain Mapping*. 2012. url: https://brainvisa.info/web/morphologist.html (visited on 03/28/2024).

[120] Bruce Fischl, Martin I. Sereno, and Anders M. Dale. "Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System". In: *NeuroImage* 9.2 (Feb. 1999), pp. 195–207. issn: 1053-8119. doi: 10.1006/nimg.1998.0396. url: https://www.sciencedirect.com/science/article/pii/S1053811998903962 (visited on 01/18/2024).

[121] Jean-Philippe Fortin, Drew Parker, Birkan Tunç, Takanori Watanabe, Mark A. Elliott, Kosha Ruparel, David R. Roalf, Theodore D. Satterthwaite, Ruben C. Gur, Raquel E. Gur, Robert T. Schultz, Ragini Verma, and Russell T. Shinohara. "Harmonization of multi-site diffusion tensor imaging data". In: *NeuroImage* 161 (Nov. 2017), pp. 149–170. issn: 1053-8119. doi: 10 . 1016 / j . neuroimage . 2017 . 08 . 047. url: https : / / www . sciencedirect . com / science / article / pii / S1053811917306948 (visited on 01/26/2024).

[122] Cynthia H Y Fu and Sergi G Costafreda. "Neuroimaging-Based Biomarkers in Psychiatry: Clinical Opportunities of a Paradigm Shift". en. In: *The Canadian Journal of Psychiatry* 58.9 (Sept. 2013). Publisher: SAGE Publications Inc, pp. 499–508. issn: 0706-7437. doi: 10 . 1177 / 070674371305800904. url: https : / / doi . org / 10 . 1177 / 070674371305800904 (visited on 04/05/2024).

[123] Kunihiko Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". en. In: *Biological Cybernetics* 36.4 (Apr. 1980), pp. 193–202. issn: 1432-0770. doi: 10.1007/BF00344251. url: https://doi.org/10.1007/BF00344251 (visited on 03/01/2024).

[124] Anna E. Fürtjes, James H. Cole, Baptiste Couvy-Duchesne, and Stuart J. Ritchie. "A quantified comparison of cortical atlases on the basis of trait morphometricity". In: *Cortex* 158 (Jan. 2023), pp. 110–126. issn: 0010-9452. doi: 10.1016/j.cortex.2022.11.001. url: https://www.sciencedirect.com/science/article/pii/S0010945222003008 (visited on 04/15/2024).

[125] Paolo Fusar-Poli, Marco Solmi, Natascia Brondino, Cathy Davies, Chungil Chae, Pierluigi Politi, Stefan Borgwardt, Stephen M. Lawrie, Josef Parnas, and Philip McGuire. "Transdiagnostic psychiatry: a systematic review". en. In: *World Psychiatry* 18.2 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wps.20631, pp. 192–207. issn: 2051-5545. doi: 10.1002/wps.20631. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/wps.20631 (visited on 01/31/2024).

[126] Yarin Gal. "Uncertainty in Deep Learning". en. PhD thesis. UNiversity of Cambridge, 2016.

[127] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". en. In: *Proceedings of The 33rd International Conference on Machine Learning*. ISSN: 1938-7228. PMLR, June 2016, pp. 1050–1059. url: https://proceedings.mlr.press/v48/gal16.html (visited on 01/23/2024).

[128] Jing Gao, Peng Li, Zhikui Chen, and Jianing Zhang. "A Survey on Deep Learning for Multimodal Data Fusion". In: *Neural Computation* 32.5 (May 2020), pp. 829–864. issn: 0899-7667. doi: 10.1162/neco_a_01273. url: https://doi.org/10.1162/neco_a_01273 (visited on 03/24/2021).

[129] Miguel Ángel García-Cabezas, Basilis Zikopoulos, and Helen Barbas. "The Structural Model: a theory linking connections, plasticity, pathology, development and evolution of the cerebral cortex". en. In: *Brain Structure and Function* 224.3 (Apr. 2019), pp. 985–1008. issn: 1863-2661. doi: 10.1007/s00429-019-01841-9. url: https://doi.org/10.1007/s00429-019-01841-9 (visited on 04/15/2024).

[130] Quentin Garrido, Randall Balestriero, Laurent Najman, and Yann Lecun. "RankMe: Assessing the Downstream Performance of Pretrained Self-Supervised Representations by Their Rank". en. In: *Proceedings of the 40th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2023, pp. 10929–10974. url: https://proceedings.mlr.press/v202/garrido23a.html (visited on 09/13/2023).

[131] Christian Gaser, Robert Dahnke, Paul M. Thompson, Florian Kurth, Eileen Luders, and Alzheimer's Disease Neuroimaging Initiative. *CAT – A Computational Anatomy Toolbox for the Analysis of Structural MRI Data*. en. Pages: 2022.06.11.495736 Section: New Results. 2016. doi: 10.1101/2022.06.11.495736. url: https://www.biorxiv.org/content/10.1101/2022.06.11.495736v2 (visited on 04/23/2024).

[132] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. *A Survey of Uncertainty in Deep Neural Networks*. arXiv:2107.03342 [cs, stat]. Jan. 2022. doi: 10.48550/arXiv.2107.03342. url: http://arxiv.org/abs/2107.03342 (visited on 03/20/2024).

[133] Sayan Ghosal, Qiang Chen, Giulio Pergola, Aaron L. Goldman, William Ulrich, Daniel R. Weinberger, and Archana Venkataraman. "A Biologically Interpretable Graph Convolutional Network to Link Genetic Risk Pathways and Imaging Phenotypes of Disease". en. In: *The Tenth International Conference on Learning Representations, {ICLR} 2022, Virtual Event, April 25-29, 2022*. Oct. 2021. url: https://openreview.net/forum?id=Lwr8We4MIxn (visited on 02/19/2024).

[134] Arna Ghosh, Arnab Kumar Mondal, Kumar Krishna Agrawal, and Blake Richards. *Investigating Power laws in Deep Representation Learning*. arXiv:2202.05808 [cs, q-bio]. Feb. 2022. doi: 10.48550/arXiv.2202.05808. url: http://arxiv.org/abs/2202.05808 (visited on 03/07/2024).

[135] Arnaud Gloaguen, Cathy Philippe, Vincent Frouin, Giulia Gennari, Ghislaine Dehaene-Lambertz, Laurent Le Brusquet, and Arthur Tenenhaus. "Multiway generalized canonical correlation analysis". In: *Biostatistics* 23.1 (Jan. 2022), pp. 240–256. issn: 1465-4644. doi: 10.1093/biostatistics/kxaa010. url: https://doi.org/10.1093/biostatistics/kxaa010 (visited on 04/17/2024).

[136] Ben Glocker, Robert Robinson, Daniel C. Castro, Qi Dou, and Ender Konukoglu. *Machine Learning with Multi-Site Imaging Data: An Empirical Study on the Impact of Scanner Effects*. arXiv:1910.04597 [cs, eess, q-bio]. Oct. 2019. doi: 10.48550/arXiv.1910.04597. url: http://arxiv.org/abs/1910.04597 (visited on 02/19/2024).

[137] Qiyong Gong, Xinyu Hu, William Pettersson-Yeo, Xin Xu, Su Lui, Nicolas Crossley, Min Wu, Hongyan Zhu, and Andrea Mechelli. "Network-Level Dysconnectivity in Drug-Naïve First-Episode Psychosis: Dissociating Transdiagnostic and Diagnosis-Specific Alterations". en. In: *Neuropsychopharmacology* 42.4 (Mar. 2017). Publisher: Nature Publishing Group, pp. 933–940. issn: 1740-634X. doi: 10.1038/npp.2016.247. url: https://www.nature.com/articles/npp2016247 (visited on 04/11/2024).

[138] Ian Goodfellow, Aaron Courville, and Yoshua Bengio. *Deep Learning*. en-US. MIT Press, 2016. url: https://mitpress.mit.edu/9780262035613/deep-learning/ (visited on 03/19/2024).

[139] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc., 2014. url: https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html (visited on 03/01/2024).

[140] Madeleine Goodkind, Simon B. Eickhoff, Desmond J. Oathes, Ying Jiang, Andrew Chang, Laura B. Jones-Hagata, Brissa N. Ortega, Yevgeniya V. Zaiko, Erika L. Roach, Mayuresh S. Korgaonkar, Stuart M. Grieve, Isaac Galatzer-Levy, Peter T. Fox, and Amit Etkin. "Identification of a Common Neurobiological Substrate for Mental Illness". In: *JAMA Psychiatry* 72.4 (Apr. 2015), pp. 305–315. issn: 2168-622X. doi: 10.1001/jamapsychiatry.2014.2206. url: https://doi.org/10.1001/jamapsychiatry.2014.2206 (visited on 03/12/2024).

[141] Robert Goodman. "The Strengths and Difficulties Questionnaire: A Research Note". en. In: *Journal of Child Psychology and Psychiatry* 38.5 (1997). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-7610.1997.tb01545.x, pp. 581–586. issn: 1469-7610. doi: 10.1111/j.1469-7610.1997.tb01545.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7610.1997.tb01545.x (visited on 04/09/2024).

[142] Karthik Gopinath, Christian Desrosiers, and Herve Lombaert. "Adaptive Graph Convolution Pooling for Brain Surface Analysis". en. In: *Information Processing in Medical Imaging*. Ed. by Albert C. S. Chung, James C. Gee, Paul A. Yushkevich, and Siqi Bao. Cham: Springer International Publishing, 2019, pp. 86–98. isbn: 978-3-030-20351-1. doi: 10.1007/978-3-030-20351-1_7.

[143] Karthik Gopinath, Christian Desrosiers, and Herve Lombaert. "Graph Convolutions on Spectral Embeddings for Cortical Surface Parcellation". In: *Medical Image Analysis* 54 (May 2019), pp. 297–305. issn: 1361-8415. doi: 10.1016/j.media.2019.03.012. url: https://www.sciencedirect.com/science/article/pii/S1361841518305243 (visited on 04/03/2024).

[144] Anirudh Goyal and Yoshua Bengio. *Inductive Biases for Deep Learning of Higher-Level Cognition*. arXiv:2011.15091 [cs, stat]. Aug. 2022. doi: 10.48550/arXiv.2011.15091. url: http://arxiv.org/abs/2011.15091 (visited on 03/05/2024).

[145] Alex Graves, Douglas Eck, Nicole Beringer, and Juergen Schmidhuber. "Biologically Plausible Speech Recognition with LSTM Neural Nets". en. In: *Biologically Inspired Approaches to Advanced Information Technology*. Ed. by Auke Jan Ijspeert, Masayuki Murata, and Naoki Wakamiya. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004, pp. 127–136. isbn: 978-3-540-27835-1. doi: 10.1007/978-3-540-27835-1_10.

[146] Douglas N. Greve, Lise Van der Haegen, Qing Cai, Steven Stufflebeam, Mert R. Sabuncu, Bruce Fischl, and Marc Brysbaert. "A Surface-based Analysis of Language Lateralization and Cortical Asymmetry". In: *Journal of Cognitive Neuroscience* 25.9 (Sept. 2013), pp. 1477–1492. issn: 0898-929X. doi: 10.1162/jocn_a_00405. url: https://doi.org/10.1162/jocn_a_00405 (visited on 01/26/2024).

[147] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, koray kavukcuoglu koray, Remi Munos, and Michal Valko. "Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 21271–21284. url: https://papers.nips.cc/paper/2020/hash/f3ada80d5c4ee70142b17b8192b2958e-Abstract.html (visited on 01/18/2024).

[148] Katherine A. Grisanzio, Andrea N. Goldstein-Piekarski, Michelle Yuyun Wang, Abdullah P. Rashed Ahmed, Zoe Samara, and Leanne M. Williams. "Transdiagnostic Symptom Clusters and Associations With Brain, Behavior, and Daily Function in Mood, Anxiety, and Trauma Disorders". In: *JAMA Psychiatry* 75.2 (Feb. 2018), pp. 201–209. issn: 2168-622X. doi: 10.1001/jamapsychiatry.2017.3951. url: https://doi.org/10.1001/jamapsychiatry.2017.3951 (visited on 03/12/2024).

[149] Logan Grosenick, Brad Klingenberg, Kiefer Katovich, Brian Knutson, and Jonathan E. Taylor. "Interpretable whole-brain prediction analysis with GraphNet". In: *NeuroImage* 72 (May 2013), pp. 304–321. issn: 1053-8119. doi: 10.1016/j.neuroimage.2012.12.062. url: https://www.sciencedirect.com/science/article/pii/S1053811912012487 (visited on 03/26/2024).

[150] N. Guigui, C. Philippe, A. Gloaguen, S. Karkar, V. Guillemot, T. Löfstedt, and V. Frouin. "Network Regularization in Imaging Genetics Improves Prediction Performances and Model Interpretability on Alzheimer's Disease". In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. ISSN: 1945-8452. Apr. 2019, pp. 1403–1406. doi: 10.1109/ISBI.2019.8759593. url: https://ieeexplore.ieee.org/document/8759593 (visited on 04/17/2024).

[151] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. "On Calibration of Modern Neural Networks". en. In: *Proceedings of the 34th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2017, pp. 1321–1330. url: https://proceedings.mlr.press/v70/guo17a.html (visited on 03/20/2024).

[152] Arpana Gupta, Emeran A. Mayer, Claudia P. Sanmiguel, John D. Van Horn, Davis Woodworth, Benjamin M. Ellingson, Connor Fling, Aubrey Love, Kirsten Tillisch, and Jennifer S. Labus. "Patterns of brain structural connectivity differentiate normal weight from overweight subjects". In: *NeuroImage: Clinical* 7 (Jan. 2015), pp. 506–517. issn: 2213-1582. doi: 10.1016/j.nicl.2015.01.005. url: https://www.sciencedirect.com/science/article/pii/S2213158215000066 (visited on 03/21/2024).

[153] Prashnna K. Gyawali, Xiaoxia Liu, James Zou, and Zihuai He. "Ensembling improves stability and power of feature selection for deep learning models". en. In: *Proceedings of the 17th Machine Learning in Computational Biology meeting*. ISSN: 2640-3498. PMLR, Dec. 2022, pp. 33–45. url: https://proceedings.mlr.press/v200/gyawali22a.html (visited on 03/20/2024).

[154] David R. Hardoon, Sandor Szedmak, and John Shawe-Taylor. "Canonical Correlation Analysis: An Overview with Application to Learning Methods". In: *Neural Computation* 16.12 (Dec. 2004), pp. 2639–2664. issn: 0899-7667. doi: 10.1162/0899766042321814. url: https://doi.org/10.1162/0899766042321814 (visited on 03/14/2024).

[155] Moritz Hardt, Ben Recht, and Yoram Singer. "Train faster, generalize better: Stability of stochastic gradient descent". en. In: *Proceedings of The 33rd International Conference on Machine Learning*. ISSN: 1938-7228. PMLR, June 2016, pp. 1225–1234. url: https://proceedings.mlr.press/v48/hardt16.html (visited on 02/29/2024).

[156] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer, 2009. isbn: 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7. url: http://link.springer.com/10.1007/978-0-387-84858-7 (visited on 02/06/2024).

[157] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. "Masked Autoencoders Are Scalable Vision Learners". en. In: *{IEEE/CVF} Conference on Computer Vision and Pattern Recognition, {CVPR} 2022, New Orleans, LA, USA, June 18-24, 2022*. 2022, pp. 16000–16009. url: https://openaccess.thecvf.com/content/CVPR2022/html/He_Masked_Autoencoders_Are_Scalable_Vision_Learners_CVPR_2022_paper (visited on 03/06/2024).

[158] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. "Momentum Contrast for Unsupervised Visual Representation Learning". en. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, June 2020, pp. 9726–9735. isbn: 978-1-72817-168-5. doi: 10.1109/CVPR42600.2020.00975. url: https://ieeexplore.ieee.org/document/9157636/ (visited on 01/18/2024).

[159] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". English. In: *2016 {IEEE} Conference on Computer Vision and Pattern Recognition, {CVPR} 2016, Las Vegas, NV, USA, June 27-30, 2016*. ISSN: 1063-6919. IEEE Computer Society, June 2016, pp. 770–778. isbn: 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90. url: https://www.computer.org/csdl/proceedings-article/cvpr/2016/8851a770/12OmNxvwoXv (visited on 03/01/2024).

[160] Ning He, Fei Li, Yuanyuan Li, Lanting Guo, Lizhou Chen, Xiaoqi Huang, Su Lui, and Qiyong Gong. "Neuroanatomical deficits correlate with executive dysfunction in boys with attention deficit hyperactivity disorder". In: *Neuroscience Letters* 600 (2015), pp. 45–49. issn: 0304-3940. doi: https://doi.org/10.1016/j.neulet.2015.05.062. url: https://www.sciencedirect.com/science/article/pii/S0304394015004292.

[161] Markus Helmer, Shaun Warrington, Ali-Reza Mohammadi-Nejad, Jie Lisa Ji, Amber Howell, Benjamin Rosand, Alan Anticevic, Stamatios N. Sotiropoulos, and John D. Murray. "On the stability of canonical correlation analysis and partial least squares with application to brain-behavior associations". en. In: *Communications Biology* 7.1 (Feb. 2024). Publisher: Nature Publishing Group, pp. 1–15. issn: 2399-3642. doi: 10.1038/

s42003-024-05869-4. url: https://www.nature.com/articles/s42003-024-05869-4 (visited on 03/21/2024).

[162] Leonie Henschel, Sailesh Conjeti, Santiago Estrada, Kersten Diers, Bruce Fischl, and Martin Reuter. "FastSurfer - A fast and accurate deep learning based neuroimaging pipeline". In: *NeuroImage* 219 (Oct. 2020), p. 117012. issn: 1053-8119. doi: 10.1016/j.neuroimage.2020.117012. url: https://www.sciencedirect.com/science/article/pii/S1053811920304985 (visited on 02/19/2024).

[163] Moritz Herrmann, Philipp Probst, Roman Hornung, Vindi Jurinovic, and Anne-Laure Boulesteix. "Large-scale benchmark study of survival prediction methods using multi-omics data". In: *Briefings in Bioinformatics* 22.3 (May 2021), bbaa167. issn: 1477-4054. doi: 10.1093/bib/bbaa167. url: https://doi.org/10.1093/bib/bbaa167 (visited on 03/26/2024).

[164] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". en. In: *5th International Conference on Learning Representations, {ICLR} 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. Nov. 2016. url: https://openreview.net/forum?id=Sy2fzU9gl (visited on 05/31/2021).

[165] G. E. Hinton and R. R. Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". In: *Science* 313.5786 (July 2006). Publisher: American Association for the Advancement of Science, pp. 504–507. doi: 10.1126/science.1127647. url: https://www.science.org/doi/10.1126/science.1127647 (visited on 03/19/2024).

[166] Geoffrey E Hinton and Richard Zemel. "Autoencoders, Minimum Description Length and Helmholtz Free Energy". In: *Advances in Neural Information Processing Systems*. Vol. 6. Morgan-Kaufmann, 1993. url: https://proceedings.neurips.cc/paper/1993/hash/9e3cfc48eccf81a0d57663e129aef3cb-Abstract.html (visited on 03/06/2024).

[167] Geoffrey E. Hinton, Alex Krizhevsky, and Sida D. Wang. "Transforming Auto-Encoders". en. In: *Artificial Neural Networks and Machine Learning – ICANN 2011*. Ed. by Timo Honkela, Włodzisław Duch, Mark Girolami, and Samuel Kaski. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2011, pp. 44–51. isbn: 978-3-642-21735-7. doi: 10.1007/978-3-642-21735-7_6.

[168] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising Diffusion Probabilistic Models". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851. url: https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html (visited on 03/01/2024).

[169] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. issn: 0899-7667. doi: 10.1162/neco.1997.9.8.1735. url: https://doi.org/10.1162/neco.1997.9.8.1735 (visited on 03/04/2024).

[170] G. Hommel. "A stagewise rejective multiple test procedure based on a modified Bonferroni test". In: *Biometrika* 75.2 (June 1988), pp. 383–386. issn: 0006-3444. doi: 10.1093/biomet/75.2.383. url: https://doi.org/10.1093/biomet/75.2.383 (visited on 03/20/2024).

[171] Andrew Hoopes, Juan Eugenio Iglesias, Bruce Fischl, Douglas Greve, and Adrian V. Dalca. "TopoFit: Rapid Reconstruction of Topologically-Correct Cortical Surfaces". en. In: *Medical Imaging with Deep Learning*. Dec. 2021. url: https://openreview.net/forum?id=-JiHeZNDY3a (visited on 04/03/2024).

[172] J J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." en. In: *Proceedings of the National Academy of Sciences* 79.8 (Apr. 1982), pp. 2554–2558. issn: 0027-8424, 1091-6490. doi: 10.1073/pnas.79.8.2554. url: https://pnas.org/doi/full/10.1073/pnas.79.8.2554 (visited on 03/08/2024).

[173] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. issn: 0893-6080. doi: 10.1016/0893-6080(89)90020-8. url: https://www.sciencedirect.com/science/article/pii/0893608089900208 (visited on 02/29/2024).

[174] Harold Hotelling. "Relations Between Two Sets of Variates". In: *Biometrika* 28.3/4 (1936). Publisher: [Oxford University Press, Biometrika Trust], pp. 321–377. issn: 0006-3444. doi: 10.2307/2333955. url: https://www.jstor.org/stable/2333955 (visited on 03/13/2024).

[175] Fengling Hu, Andrew A. Chen, Hannah Horng, Vishnu Bashyam, Christos Davatzikos, Aaron Alexander-Bloch, Mingyao Li, Haochang Shou, Theodore D. Satterthwaite, Meichen Yu, and Russell T. Shinohara. "Image harmonization: A review of statistical and deep learning methods for removing batch effects and evaluation metrics for effective harmonization". In: *NeuroImage* 274 (July 2023), p. 120125. issn: 1053-8119. doi: 10.1016/j.neuroimage.2023.120125. url: https://www.sciencedirect.com/science/article/pii/S1053811923002719 (visited on 03/13/2024).

[176] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. "Densely Connected Convolutional Networks". In: *2017 {IEEE} Conference on Computer Vision and Pattern Recognition, {CVPR} 2017, Honolulu, HI, USA, July 21-26, 2017*. 2017, pp. 4700–4708. url: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html (visited on 03/01/2024).

[177] Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. "What Makes Multi-Modal Learning Better than Single (Provably)". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 10944–10956. url: https://proceedings.neurips.cc/paper/2021/hash/5aa3405a3f865c10f420a4a7b55cbff3-Abstract.html (visited on 03/28/2024).

[178] D. H. Hubel and T. N. Wiesel. “Receptive fields and functional architecture of monkey striate cortex”. en. In: *The Journal of Physiology* 195.1 (1968). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1968.sp008455, pp. 215–243. issn: 1469-7793. doi: 10.1113/jphysiol.1968.sp008455. url: https://onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1968.sp008455 (visited on 03/01/2024).

[179] Eyke Hüllermeier and Willem Waegeman. “Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods”. en. In: *Machine Learning* 110.3 (Mar. 2021), pp. 457–506. issn: 1573-0565. doi: 10.1007/s10994-021-05946-3. url: https://doi.org/10.1007/s10994-021-05946-3 (visited on 01/23/2024).

[180] Alex Ing, Philipp G. Sämann, et al. “Identification of neurobehavioural symptom groups based on shared brain mechanisms”. en. In: *Nature Human Behaviour* 3.12 (Dec. 2019). Number: 12 Publisher: Nature Publishing Group, pp. 1306–1318. issn: 2397-3374. doi: 10.1038/s41562-019-0738-8. url: https://www.nature.com/articles/s41562-019-0738-8 (visited on 09/25/2023).

[181] Thomas Insel, Bruce Cuthbert, Marjorie Garvey, Robert Heinssen, Daniel S. Pine, Kevin Quinn, Charles Sanislow, and Philip Wang. “Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders”. In: *American Journal of Psychiatry* 167.7 (July 2010). Publisher: American Psychiatric Publishing, pp. 748–751. issn: 0002-953X. doi: 10.1176/appi.ajp.2010.09091379. url: https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2010.09091379 (visited on 03/12/2024).

[182] Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. *Mixtral of Experts*. arXiv:2401.04088 [cs]. Jan. 2024. doi: 10.48550/arXiv.2401.04088. url: http://arxiv.org/abs/2401.04088 (visited on 04/02/2024).

[183] Chiyu Max Jiang, Jingwei Huang, Karthik Kashinath, Prabhat, Philip Marcus, and Matthias Niessner. “Spherical CNNs on Unstructured Grids”. en. In: *7th International Conference on Learning Representations, {ICLR} 2019, New Orleans, LA, USA, May 6-9, 2019*. Sept. 2018. url: https://openreview.net/forum?id=Bkl-43C9FQ (visited on 01/18/2024).

[184] Li Jing, Pascal Vincent, Yann LeCun, and Yuandong Tian. “Understanding Dimensional Collapse in Contrastive Self-supervised Learning”. en. In: *The Tenth International Conference on Learning Representations, {ICLR} 2022, Virtual Event, April 25-29, 2022*. Oct. 2021. url: https://openreview.net/forum?id=YevsQ05DEN7 (visited on 03/07/2024).

[185] P. B. Jones. "Adult mental health disorders and their age at onset". en. In: *The British Journal of Psychiatry* 202.s54 (Jan. 2013), s5–s10. issn: 0007-1250, 1472-1465. doi: 10.1192/bjp.bp.112.119164. url: https://www.cambridge.org/core/journals/the-british-journal-of-psychiatry/article/adult-mental-health-disorders-and-their-age-at-onset/13F1A156235E5FF0D904F2CE2FDC053F (visited on 03/13/2024).

[186] Gargi Joshi, Rahee Walambe, and Ketan Kotecha. "A Review on Explainability in Multimodal Deep Neural Nets". en. In: *IEEE Access* 9 (2021), pp. 59800–59821. issn: 2169-3536. doi: 10.1109/ACCESS.2021.3070212. url: https://ieeexplore.ieee.org/document/9391727/ (visited on 03/15/2024).

[187] Tom Joy, Sebastian Schmon, Philip Torr, Siddharth N, and Tom Rainforth. "Capturing Label Characteristics in VAEs". en. In: *9th International Conference on Learning Representations, {ICLR} 2021, Virtual Event, Austria, May 3-7, 2021*. Sept. 2020. url: https://openreview.net/forum?id=wQRlSUZ5V7B (visited on 09/08/2021).

[188] Tom Joy, Yuge Shi, Philip Torr, Tom Rainforth, Sebastian M. Schmon, and Siddharth N. "Learning Multimodal VAEs through Mutual Supervision". en. In: *The Tenth International Conference on Learning Representations, {ICLR} 2022, Virtual Event, April 25-29, 2022*. Oct. 2021. url: https://openreview.net/forum?id=1xXvPrAshao (visited on 07/12/2023).

[189] Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. *Exploring the Limits of Language Modeling*. arXiv:1602.02410 [cs]. Feb. 2016. doi: 10.48550/arXiv.1602.02410. url: http://arxiv.org/abs/1602.02410 (visited on 03/04/2024).

[190] Nicole R. Karcher and Deanna M. Barch. "The ABCD study: understanding the development of risk for mental and physical health outcomes". en. In: *Neuropsychopharmacology* 46.1 (Jan. 2021). Number: 1 Publisher: Nature Publishing Group, pp. 131–142. issn: 1740-634X. doi: 10.1038/s41386-020-0736-6. url: https://www.nature.com/articles/s41386-020-0736-6 (visited on 02/19/2024).

[191] Guanzhou Ke, Yang Yu, Guoqing Chao, Xiaoli Wang, Chenyang Xu, and Shengfeng He. "Disentangling Multi-view Representations Beyond Inductive Bias". In: *Proceedings of the 31st ACM International Conference on Multimedia*. arXiv:2308.01634 [cs]. Oct. 2023, pp. 2582–2590. doi: 10.1145/3581783.3611794. url: http://arxiv.org/abs/2308.01634 (visited on 02/23/2024).

[192] Alex Kendall and Yarin Gal. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. url: https://papers.nips.cc/paper_files/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html (visited on 01/23/2024).

[193] Matcheri S. Keshavan, David W. Morris, John A Sweeney, Godfrey Pearlson, Gunvant Thaker, Larry J Seidman, Shaun M. Eack, and Carol Tamminga. "A dimensional approach to the psychosis spectrum between bipolar disorder and schizophrenia: The Schizo-Bipolar Scale". In: *Schizophrenia Research* 133.1-3 (Dec. 2011), pp. 250–254. issn: 0920-9964. doi: 10.1016/j.schres.2011.09.005. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3381911/ (visited on 04/08/2024).

[194] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. "Supervised Contrastive Learning". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 18661–18673. url: https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html (visited on 01/18/2024).

[195] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *3rd International Conference on Learning Representations, {ICLR} 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings}*. arXiv:1412.6980 [cs]. San Diego, CA, USA: arXiv, 2015. doi: 10.48550/arXiv.1412.6980. url: http://arxiv.org/abs/1412.6980 (visited on 03/07/2024).

[196] Diederik P. Kingma, Danilo J. Rezende, Shakir Mohamed, and Max Welling. "Semi-Supervised Learning with Deep Generative Models". In: *arXiv:1406.5298 [cs, stat]* (Oct. 2014). arXiv: 1406.5298. url: http://arxiv.org/abs/1406.5298 (visited on 02/15/2021).

[197] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". en. In: *2nd International Conference on Learning Representations, {ICLR} 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. Dec. 2013. url: https://openreview.net/forum?id=33X9fd2-9FyZd (visited on 03/19/2024).

[198] Thomas N. Kipf and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks". In: *arXiv:1609.02907 [cs, stat]* (Feb. 2017). arXiv: 1609.02907. url: http://arxiv.org/abs/1609.02907 (visited on 12/15/2020).

[199] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. "1D convolutional neural networks and applications: A survey". In: *Mechanical Systems and Signal Processing* 151 (Apr. 2021), p. 107398. issn: 0888-3270. doi: 10.1016/j.ymssp.2020.107398. url: https://www.sciencedirect.com/science/article/pii/S0888327020307846 (visited on 03/01/2024).

[200] Ryan Kiros, Ruslan Salakhutdinov, and Richard S. Zemel. *Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models*. arXiv:1411.2539 [cs]. Nov. 2014. doi: 10.48550/arXiv.1411.2539. url: http://arxiv.org/abs/1411.2539 (visited on 03/04/2024).

[201] Armen Der Kiureghian and Ove Ditlevsen. "Aleatory or epistemic? Does it matter?" In: *Structural Safety*. Risk Acceptance and Risk Communication 31.2 (Mar. 2009), pp. 105–112. issn: 0167-4730. doi: 10.1016/j.strusafe.2008.06.020. url: https://www.sciencedirect.com/science/article/pii/S0167473008000556 (visited on 01/23/2024).

[202] Arto Klami and Samuel Kaski. "Local dependent components". In: *Proceedings of the 24th international conference on Machine learning*. ICML '07. New York, NY, USA: Association for Computing Machinery, June 2007, pp. 425–432. isbn: 978-1-59593-793-3. doi: 10.1145/1273496.1273550. url: https://doi.org/10.1145/1273496.1273550 (visited on 03/14/2024).

[203] Eduard T. Klapwijk, Wouter van den Bos, Christian K. Tamnes, Nora M. Raschle, and Kathryn L. Mills. "Opportunities for increased reproducibility and replicability of developmental neuroimaging". In: *Developmental Cognitive Neuroscience* 47 (Feb. 2021), p. 100902. issn: 1878-9293. doi: 10.1016/j.dcn.2020.100902. url: https://www.sciencedirect.com/science/article/pii/S1878929320301511 (visited on 04/10/2024).

[204] Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. "Multimodal machine learning in precision health: A scoping review". en. In: *npj Digital Medicine* 5.1 (Nov. 2022). Publisher: Nature Publishing Group, pp. 1–14. issn: 2398-6352. doi: 10.1038/s41746-022-00712-8. url: https://www.nature.com/articles/s41746-022-00712-8 (visited on 03/27/2024).

[205] Mark A Kramer. "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks". en. In: *AIChE Journal* 37.2 (1991).

[206] Nikolaus Kriegeskorte, Marieke Mur, and Peter A. Bandettini. "Representational similarity analysis - connecting the branches of systems neuroscience". English. In: *Frontiers in Systems Neuroscience* 2 (Nov. 2008). Publisher: Frontiers. issn: 1662-5137. doi: 10.3389/neuro.06.004.2008. url: https://www.frontiersin.org/articles/10.3389/neuro.06.004.2008 (visited on 03/20/2024).

[207] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc., 2012. url: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (visited on 02/29/2024).

[208] Sayantan Kumar, Philip Payne, and Aristeidis Sotiras. "mmNormVAE: Normative Modeling on Multimodal Neuroimaging Data using Variational Autoencoders". en. In: Oct. 2023. url: https://openreview.net/forum?id=khB5CQidql (visited on 04/18/2024).

[209] Sayantan Kumar, Philip R. O. Payne, and Aristeidis Sotiras. "Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in Alzheimer's disease". In: *Medical Imaging 2023: Computer-Aided Diagnosis*. Vol. 12465. SPIE, Apr. 2023, p. 1246503. doi: 10.1117/12.2654369. url: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12465/1246503/Normative-modeling-using-multimodal-variational-autoencoders-to-identify-abnormal-brain/10.1117/12.2654369.full (visited on 04/18/2024).

[210] Jennifer S. Labus, John D. Van Horn, Arpana Gupta, Mher Alaverdyan, Carinna Torgerson, Cody Ashe-McNalley, Andrei Irimia, Jui-Yang Hong, Bruce Naliboff, Kirsten Tillisch, and Emeran A. Mayer. "Multivariate morphological brain signatures predict patients with chronic abdominal pain from healthy control subjects". en-US. In: *PAIN* 156.8 (Aug. 2015), p. 1545. issn: 0304-3959. doi: 10.1097/j.pain.0000000000000196. url: https://journals.lww.com/pain/abstract/2015/08000/multivariate_morphological_brain_signatures.22.aspx (visited on 03/21/2024).

[211] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. url: https://proceedings.neurips.cc/paper_files/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html (visited on 01/22/2024).

[212] Rodney LaLonde and Ulas Bagci. "Capsules for Object Segmentation". en. In: *Medical Imaging with Deep Learning*. 2018. url: https://openreview.net/forum?id=B1L4Gb3sf (visited on 03/19/2024).

[213] Ana Lawry Aguila, James Chapman, and Andre Altmann. "Multi-modal Variational Autoencoders for Normative Modelling Across Multiple Imaging Modalities". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. Ed. by Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor. Cham: Springer Nature Switzerland, 2023, pp. 425–434. isbn: 978-3-031-43907-0. doi: 10.1007/978-3-031-43907-0_41.

[214] Ana Lawry Aguila, James Chapman, Mohammed Janahi, and Andre Altmann. "Conditional VAEs for Confound Removal and Normative Modelling of Neurodegenerative Diseases". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Ed. by Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li. Cham: Springer Nature Switzerland, 2022, pp. 430–440. isbn: 978-3-031-16431-6. doi: 10.1007/978-3-031-16431-6_41.

[215] Kim-Anh Lê Cao, Simon Boitard, and Philippe Besse. "Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems". In: *BMC Bioinformatics* 12.1 (June 2011), p. 253. issn: 1471-2105. doi: 10.1186/1471-2105-12-253. url: https://doi.org/10.1186/1471-2105-12-253 (visited on 03/21/2024).

[216] Edith Le Floch, Vincent Guillemot, Vincent Frouin, Philippe Pinel, Christophe Lalanne, Laura Trinchera, Arthur Tenenhaus, Antonio Moreno, Monica Zilbovicius, Thomas Bourgeron, Stanislas Dehaene, Bertrand Thirion, Jean-Baptiste Poline, and Edouard Duchesnay. "Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse Partial Least Squares". In: *NeuroImage* 63.1 (Oct. 2012), pp. 11–24. issn: 1053-8119. doi: 10.1016/j.neuroimage.2012.06.061. url: https://www.sciencedirect.com/science/article/pii/S1053811912006775 (visited on 03/21/2024).

[217] Yann Le Guen, Guillaume Auzias, François Leroy, Marion Noulhiane, Ghislaine Dehaene-Lambertz, Edouard Duchesnay, Jean-François Mangin, Olivier Coulon, and Vincent Frouin. "Genetic Influence on the Sulcal Pits: On the Origin of the First Cortical Folds". In: *Cerebral Cortex* 28.6 (June 2018), pp. 1922–1933. issn: 1047-3211. doi: 10.1093/cercor/bhx098. url: https://doi.org/10.1093/cercor/bhx098 (visited on 04/07/2024).

[218] Leo Lebrat, Rodrigo Santa Cruz, Frederic de Gournay, Darren Fu, Pierrick Bourgeat, Jurgen Fripp, Clinton Fookes, and Olivier Salvado. "CorticalFlow: A Diffeomorphic Mesh Transformer Network for Cortical Surface Reconstruction". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 29491–29505. url: https://papers.nips.cc/paper_files/paper/2021/hash/f6b5f8c32c65fee991049a55dc97d1ce-Abstract.html (visited on 04/03/2024).

[219] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (Nov. 1998). Conference Name: Proceedings of the IEEE, pp. 2278–2324. issn: 1558-2256. doi: 10.1109/5.726791. url: https://ieeexplore.ieee.org/document/726791 (visited on 02/29/2024).

[220] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". en. In: *Nature* 521.7553 (May 2015). Number: 7553 Publisher: Nature Publishing Group, pp. 436–444. issn: 1476-4687. doi: 10.1038/nature14539. url: https://www.nature.com/articles/nature14539 (visited on 02/29/2024).

[221] Changhee Lee and Mihaela van der Schaar. "A Variational Information Bottleneck Approach to Multi-Omics Data Integration". en. In: *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*. ISSN: 2640-3498. PMLR, Mar. 2021, pp. 1513–1521. url: https://proceedings.mlr.press/v130/lee21a.html (visited on 03/26/2024).

[222] Mihee Lee and Vladimir Pavlovic. "Private-Shared Disentangled Multimodal VAE for Learning of Latent Representations". en. In: *{IEEE} Conference on Computer Vision and Pattern Recognition Workshops, {CVPR} Workshops 2021, virtual, June 19-25, 2021*. 2021, pp. 1692–1700. url: https://openaccess.thecvf.com/content/CVPR2021W/MULA/html/Lee_Private-Shared_Disentangled_Multimodal_VAE_for_Learning_of_Latent_Representations_CVPRW_2021_paper.html (visited on 07/12/2023).

[223] Jason P. Lerch, André J. W. van der Kouwe, Armin Raznahan, Tomáš Paus, Heidi Johansen-Berg, Karla L. Miller, Stephen M. Smith, Bruce Fischl, and Stamatios N. Sotiropoulos. "Studying neuroanatomy using MRI". en. In: *Nature Neuroscience* 20.3 (Mar. 2017). Publisher: Nature Publishing Group, pp. 314–326. issn: 1546-1726. doi: 10.1038/nn.4501. url: https://www.nature.com/articles/nn.4501 (visited on 04/07/2024).

[224] Caiyan Li and Hongzhe Li. "Network-constrained regularization and variable selection for analysis of genomic data". In: *Bioinformatics* 24.9 (May 2008), pp. 1175–1182. issn: 1367-4803. doi: 10.1093/bioinformatics/btn081. url: https://doi.org/10.1093/bioinformatics/btn081 (visited on 03/26/2024).

[225] Xinwei Li, Jia Tan, Panyu Wang, Hong Liu, Zhangyong Li, and Wei Wang. "Anatomically constrained squeeze-and-excitation graph attention network for cortical surface parcellation". In: *Computers in Biology and Medicine* 140 (Jan. 2022), p. 105113. issn: 0010-4825. doi: 10.1016/j.compbiomed.2021.105113. url: https://www.sciencedirect.com/science/article/pii/S0010482521009070 (visited on 04/03/2024).

[226] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. "Explainable AI: A Review of Machine Learning Interpretability Methods". en. In: *Entropy* 23.1 (Jan. 2021). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 18. issn: 1099-4300. doi: 10.3390/e23010018. url: https://www.mdpi.com/1099-4300/23/1/18 (visited on 03/05/2024).

[227] Seppo Linnainmaa. "Taylor expansion of the accumulated rounding error". en. In: *BIT Numerical Mathematics* 16.2 (June 1976), pp. 146–160. issn: 1572-9125. doi: 10.1007/BF01931367. url: https://doi.org/10.1007/BF01931367 (visited on 02/29/2024).

[228] Zachary C. Lipton. "The mythos of model interpretability". In: *Communications of the ACM* 61.10 (Sept. 2018), pp. 36–43. issn: 0001-0782. doi: 10.1145/3233231. url: https://dl.acm.org/doi/10.1145/3233231 (visited on 03/05/2024).

[229] Jianfang Liu, Tara Lichtenberg, et al. "An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics". In: *Cell* 173.2 (Apr. 2018), 400–416.e11. issn: 0092-8674. doi: 10.1016/j.cell.2018.02.052. url: https://www.sciencedirect.com/science/article/pii/S0092867418302290 (visited on 03/26/2024).

[230] Mengting Liu, Ben A. Duffy, Zhe Sun, Arthur W. Toga, A. James Barkovich, Duan Xu, and Hosung Kim. "Deep Learning of Cortical Surface Features Using Graph-Convolution Predicts Neonatal Brain Age and Neurodevelopmental Outcome". In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. ISSN: 1945-8452. Apr. 2020, pp. 1335–1338. doi: 10.1109/ISBI45749.2020.9098556. url: https://ieeexplore.ieee.org/document/9098556 (visited on 04/03/2024).

[231] Peirong Liu, Zhengwang Wu, Gang Li, Pew-Thian Yap, and Dinggang Shen. "Deep Modeling of Growth Trajectories for Longitudinal Prediction of Missing Infant Cortical Surfaces". en. In: *Information Processing in Medical Imaging*. Ed. by Albert C. S. Chung, James C. Gee, Paul A. Yushkevich, and Siqi Bao. Cham: Springer International Publishing, 2019, pp. 277–288. isbn: 978-3-030-20351-1. doi: 10.1007/978-3-030-20351-1_21.

[232] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. "Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations". In: *arXiv:1811.12359 [cs, stat]* (June 2019). arXiv: 1811.12359. url: http://arxiv.org/abs/1811.12359 (visited on 04/20/2021).

[233] Steve Lukito, Luke Norman, Christina Carlisi, Joaquim Radua, Heledd Hart, Emily Simonoff, and Katya Rubia. "Comparative meta-analyses of brain structural and functional abnormalities during cognitive control in attention-deficit/hyperactivity disorder and autism spectrum disorder". en. In: *Psychological Medicine* 50.6 (Apr. 2020). Publisher: Cambridge University Press, pp. 894–919. issn: 0033-2917, 1469-8978. doi: 10.1017/S0033291720000574. url: https://www.cambridge.org/core/journals/psychological-medicine/article/abs/comparative-metaanalyses-of-brain-structural-and-functional-abnormalities-during-cognitive-control-in-attentiondeficithyperactivity-disorder-and-autism-spectrum-disorder/ED38EB66D818A8D1A531C62D8097572B (visited on 02/01/2024).

[234] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. url: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (visited on 03/06/2024).

[235] Qiang Ma, Liu Li, Emma C. Robinson, Bernhard Kainz, Daniel Rueckert, and Amir Alansary. *CortexODE: Learning Cortical Surface Reconstruction by Neural ODEs*. arXiv:2202.08329 [cs, eess]. Sept. 2022. doi: 10.48550/arXiv.2202.08329. url: http://arxiv.org/abs/2202.08329 (visited on 04/03/2024).

[236] Qiang Ma, Emma C. Robinson, Bernhard Kainz, Daniel Rueckert, and Amir Alansary. "PialNN: A Fast Deep Learning Framework for Cortical Pial Surface Reconstruction". en. In: *Machine Learning in Clinical Neuroimaging*. Ed. by Ahmed Abdulkadir, Seyed Mostafa Kia, Mohamad Habes, Vinod Kumar, Jane Maryam Rondina, Chantal Tax, and Thomas Wolfers. Cham: Springer International Publishing, 2021, pp. 73–81. isbn: 978-3-030-87586-2. doi: 10.1007/978-3-030-87586-2_8.

[237] Laurens van der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. issn: 1533-7928. url: http://jmlr.org/papers/v9/vandermaaten08a.html (visited on 03/07/2024).

[238] Annmarie MacNamara, Heide Klumpp, Amy E. Kennedy, Scott A. Langenecker, and K. Luan Phan. "Transdiagnostic neural correlates of affective face processing in anxiety and depression". en. In: *Depression and Anxiety* 34.7 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/da.22631, pp. 621–631. issn: 1520-6394. doi: 10.1002/da.22631. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/da.22631 (visited on 02/01/2024).

[239] Jean-François Mangin, Eric Jouvent, and Arnaud Cachia. “In-vivo measurement of cortical morphology: means and meanings”. en-US. In: *Current Opinion in Neurology* 23.4 (Aug. 2010), p. 359. issn: 1350-7540. doi: 10.1097/WCO.0b013e32833a0afc. url: https://journals.lww.com/co-neurology/abstract/2010/08000/in_vivo_measurement_of_cortical_morphology__means.4.aspx (visited on 04/07/2024).

[240] Scott Marek, Brenden Tervo-Clemmens, et al. “Reproducible brain-wide association studies require thousands of individuals”. en. In: *Nature* 603.7902 (Mar. 2022). Publisher: Nature Publishing Group, pp. 654–660. issn: 1476-4687. doi: 10.1038/s41586-022-04492-9. url: https://www.nature.com/articles/s41586-022-04492-9 (visited on 03/12/2024).

[241] Andre F. Marquand, Seyed Mostafa Kia, Mariam Zabihi, Thomas Wolfers, Jan K. Buitelaar, and Christian F. Beckmann. “Conceptualizing mental disorders as deviations from normative functioning”. en. In: *Molecular Psychiatry* 24.10 (Oct. 2019). Publisher: Nature Publishing Group, pp. 1415–1424. issn: 1476-5578. doi: 10.1038/s41380-019-0441-1. url: https://www.nature.com/articles/s41380-019-0441-1 (visited on 03/12/2024).

[242] Andre F. Marquand, Iead Rezek, Jan Buitelaar, and Christian F. Beckmann. “Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies”. In: *Biological Psychiatry*. Obsessive-Compulsive Disorder 80.7 (Oct. 2016), pp. 552–561. issn: 0006-3223. doi: 10.1016/j.biopsych.2015.12.023. url: https://www.sciencedirect.com/science/article/pii/S0006322316000020 (visited on 03/12/2024).

[243] Andre F. Marquand, Thomas Wolfers, Maarten Mennes, Jan Buitelaar, and Christian F. Beckmann. “Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders”. English. In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 1.5 (Sept. 2016). Publisher: Elsevier, pp. 433–447. issn: 2451-9022. doi: 10.1016/j.bpsc.2016.04.002. url: https://www.biologicalpsychiatrycnni.org/article/S2451-9022(16)30030-1/fulltext (visited on 03/12/2024).

[244] Emile Mathieu, Tom Rainforth, N. Siddharth, and Yee Whye Teh. “Disentangling Disentanglement in Variational Autoencoders”. In: *arXiv:1812.02833 [cs, stat]* (June 2019). arXiv: 1812.02833. url: http://arxiv.org/abs/1812.02833 (visited on 04/22/2021).

[245] Julia Mattu, Jeff Angwin, Lauren Larson, and Surya Kirchner. *Machine Bias*. en. May 2016. url: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (visited on 03/05/2024).

[246] Rick Mayes and Allan V. Horwitz. “DSM-III and the revolution in the classification of mental illness”. en. In: *Journal of the History of the Behavioral Sciences* 41.3 (2005). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jhbs.20103, pp. 249–267. issn: 1520-6696. doi: 10.1002/jhbs.20103. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/jhbs.20103 (visited on 03/12/2024).

[247] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. "UMAP: Uniform Manifold Approximation and Projection". en. In: *Journal of Open Source Software* 3.29 (Sept. 2018), p. 861. issn: 2475-9066. doi: 10.21105/joss.00861. url: https://joss.theoj.org/papers/10.21105/joss.00861 (visited on 03/07/2024).

[248] Roger McLendon, Allan Friedman, et al. "Comprehensive genomic characterization defines human glioblastoma genes and core pathways". en. In: *Nature* 455.7216 (Oct. 2008). Publisher: Nature Publishing Group, pp. 1061–1068. issn: 1476-4687. doi: 10.1038/nature07385. url: https://www.nature.com/articles/nature07385 (visited on 04/08/2024).

[249] Matthew A. Meier and Madeline H. Meier. "Clinical implications of a general psychopathology factor: A cognitive–behavioral transdiagnostic group treatment for community mental health". In: *Journal of Psychotherapy Integration* 28.3 (2018). Place: US Publisher: Educational Publishing Foundation, pp. 253–268. issn: 1573-3696. doi: 10.1037/int0000095.

[250] Nicolai Meinshausen and Peter Bühlmann. "Stability selection". en. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.4 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2010.00740.x, pp. 417–473. issn: 1467-9868. doi: 10.1111/j.1467-9868.2010.00740.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00740.x (visited on 01/18/2024).

[251] Charles Mellerio, Pauline Roca, Francine Chassoux, Florian Danière, Arnaud Cachia, Stéphanie Lion, Olivier Naggara, Bertrand Devaux, Jean-François Meder, and Catherine Oppenheim. "The Power Button Sign: A Newly Described Central Sulcal Pattern on Surface Rendering MR Images of Type 2 Focal Cortical Dysplasia". In: *Radiology* 274.2 (Feb. 2015). Publisher: Radiological Society of North America, pp. 500–507. issn: 0033-8419. doi: 10.1148/radiol.14140773. url: https://pubs.rsna.org/doi/abs/10.1148/radiol.14140773 (visited on 04/07/2024).

[252] Bjoern H. Menze, Andras Jakab, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". In: *IEEE Transactions on Medical Imaging* 34.10 (Oct. 2015). Conference Name: IEEE Transactions on Medical Imaging, pp. 1993–2024. issn: 1558-254X. doi: 10.1109/TMI.2014.2377694. url: https://ieeexplore.ieee.org/document/6975210 (visited on 02/19/2024).

[253] Angeline Mihailov, Cathy Philippe, Arnaud Gloaguen, Antoine Grigis, Charles Laidi, Camille Piguet, Josselin Houenou, and Vincent Frouin. "Cortical signatures in behaviorally clustered autistic traits subgroups: a population-based study". en. In: *Translational Psychiatry* 10.1 (Dec. 2020), p. 207. issn: 2158-3188. doi: 10.1038/s41398-020-00894-3. url: http://www.nature.com/articles/s41398-020-00894-3 (visited on 11/16/2020).

[254] Agoston Mihalik, James Chapman, Rick A. Adams, Nils R. Winter, Fabio S. Ferreira, John Shawe-Taylor, and Janaina Mourão-Miranda. "Canonical Correlation Analysis and Partial Least Squares for Identifying Brain–Behavior Associations: A Tutorial and a Comparative Study". In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 7.11 (Nov. 2022), pp. 1055–1067. issn: 2451-9022. doi: 10.1016/j.bpsc.2022.07.012. url: https://www.sciencedirect.com/science/article/pii/S2451902222001859 (visited on 09/15/2023).

[255] Agoston Mihalik, Fabio S. Ferreira, et al. "Multiple Holdouts With Stability: Improving the Generalizability of Machine Learning Analyses of Brain–Behavior Relationships". en. In: *Biological Psychiatry* 87.4 (Feb. 2020), pp. 368–376. issn: 00063223. doi: 10.1016/j.biopsych.2019.12.001. url: https://linkinghub.elsevier.com/retrieve/pii/S0006322319319183 (visited on 09/25/2023).

[256] Michael P. Ph D. Milham, Damien PA-C. Fair, Maarten Ph D. Mennes, and Stewart H. M. D. Mostofsky. "The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience". English. In: *Frontiers in Systems Neuroscience* 6 (Sept. 2012). Publisher: Frontiers. issn: 1662-5137. doi: 10.3389/fnsys.2012.00062. url: https://www.frontiersin.org/articles/10.3389/fnsys.2012.00062 (visited on 03/13/2024).

[257] Serge A. Mitelman. "Transdiagnostic neuroimaging in psychiatry: A review". In: *Psychiatry Research*. Celebrating the Contributions of Monte and Sherry Buchsbaum and their Impact on Psychiatry Research 277 (July 2019), pp. 23–38. issn: 0165-1781. doi: 10.1016/j.psychres.2019.01.026. url: https://www.sciencedirect.com/science/article/pii/S0165178118320845 (visited on 04/11/2024).

[258] Farida Mohsen, Hazrat Ali, Nady El Hajj, and Zubair Shah. "Artificial intelligence-based methods for fusion of electronic health records and imaging data". en. In: *Scientific Reports* 12.1 (Oct. 2022). Publisher: Nature Publishing Group, p. 17981. issn: 2045-2322. doi: 10.1038/s41598-022-22514-4. url: https://www.nature.com/articles/s41598-022-22514-4 (visited on 03/26/2024).

[259] Dmitry Molchanov, Arsenii Ashukha, and Dmitry Vetrov. "Variational Dropout Sparsifies Deep Neural Networks". In: *arXiv:1701.05369 [cs, stat]* (June 2017). arXiv: 1701.05369. url: http://arxiv.org/abs/1701.05369 (visited on 12/08/2020).

[260] João M. Monteiro, Anil Rao, John Shawe-Taylor, and Janaina Mourão-Miranda. "A multiple hold-out framework for Sparse Partial Least Squares". In: *Journal of Neuroscience Methods* 271 (Sept. 2016), pp. 182–194. issn: 0165-0270. doi: 10.1016/j.jneumeth.2016.06.011. url: https://www.sciencedirect.com/science/article/pii/S0165027016301327 (visited on 01/18/2024).

[261] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M. Bronstein. "Geometric deep learning on graphs and manifolds using mixture model CNNs". In: *arXiv:1611.08402 [cs]* (Dec. 2016). arXiv: 1611.08402. url: http://arxiv.org/abs/1611.08402 (visited on 12/15/2020).

[262] Mahmoud Mostapha, SunHyung Kim, Guorong Wu, Leo Zsembik, Stephen Pizer, and Martin Styner. "Non-Euclidean, convolutional learning on cortical brain surfaces". In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. ISSN: 1945-8452. Apr. 2018, pp. 527–530. doi: 10.1109/ISBI.2018.8363631. url: https://ieeexplore.ieee.org/document/8363631 (visited on 04/03/2024).

[263] Declan Murphy and Will Spooren. "EU-AIMS: a boost to autism research". en. In: *Nature Reviews Drug Discovery* 11.11 (Nov. 2012). Publisher: Nature Publishing Group, pp. 815–816. issn: 1474-1784. doi: 10.1038/nrd3881. url: https://www.nature.com/articles/nrd3881 (visited on 03/13/2024).

[264] Hajer Nakua, Ju-Chi Yu, Hervé Abdi, Colin Hawco, Aristotle Voineskos, Sean Hill, Meng-Chuan Lai, Anne L. Wheeler, Anthony Randal McIntosh, and Stephanie H. Ameis. *Comparing the stability and reproducibility of brain-behaviour relationships found using Canonical Correlation Analysis and Partial Least Squares within the ABCD Sample*. en. Pages: 2023.03.08.531763 Section: New Results. Mar. 2023. doi: 10.1101/2023.03.08.531763. url: https://www.biorxiv.org/content/10.1101/2023.03.08.531763v1 (visited on 03/21/2024).

[265] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. "What is being transferred in transfer learning?" In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 512–523. url: https://proceedings.neurips.cc/paper/2020/hash/0607f4c705595b911a4f3e7a127b44e0-Abstract.html (visited on 03/04/2024).

[266] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. "Multimodal deep learning". In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. ICML'11. Madison, WI, USA: Omnipress, June 2011, pp. 689–696. isbn: 978-1-4503-0619-5. (Visited on 03/19/2024).

[267] Gia H. Ngo, Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert R. Sabuncu. "From Connectomic to Task-Evoked Fingerprints: Individualized Prediction of Task Contrasts from Resting-State Functional Connectivity". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz. Cham: Springer International Publishing, 2020, pp. 62–71. isbn: 978-3-030-59728-3. doi: 10.1007/978-3-030-59728-3_7.

[268] Linn B. Norbom, Nhat Trung Doan, Dag Alnæs, Tobias Kaufmann, Torgeir Moberget, Jaroslav Rokicki, Ole A. Andreassen, Lars T. Westlye, and Christian K. Tamnes. "Probing Brain Developmental Patterns of Myelination and Associations With Psychopathology in Youths Using Gray/White Matter Contrast". In: *Biological Psychiatry*. Transdiagnostic Perspectives on Psychiatric Disorders 85.5 (Mar. 2019), pp. 389–398. issn: 0006-3223. doi: 10.1016/j.biopsych.2018.09.027. url: https://www.sciencedirect.com/science/article/pii/S0006322318319255 (visited on 04/02/2024).

[269] Brian A. Nosek, Tom E. Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S. Corker, Anna Dreber, Fiona Fidler, Joe Hilgard, Melissa Kline Struhl, Michèle B. Nuijten, Julia M. Rohrer, Felipe Romero, Anne M. Scheel, Laura D. Scherer, Felix D. Schönbrodt, and Simine Vazire. "Replicability, Robustness, and Reproducibility in Psychological Science". In: *Annual Review of Psychology* 73.1 (2022). _eprint: https://doi.org/10.1146/annurev-psych-020821-114157, pp. 719–748. doi: 10.1146/annurev-psych-020821-114157. url: https://doi.org/10.1146/annurev-psych-020821-114157 (visited on 03/12/2024).

[270] Adrian L. Oblak, Terrell T. Gibbs, and Gene J. Blatt. "Decreased GABAB receptors in the cingulate cortex and fusiform gyrus in Autism". en. In: *Journal of Neurochemistry* 114.5 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1471-4159.2010.06858.x, pp. 1414–1423. issn: 1471-4159. doi: 10.1111/j.1471-4159.2010.06858.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-4159.2010.06858.x (visited on 02/12/2024).

[271] Lennart M. Oblong, Alberto Llera, et al. "Linking functional and structural brain organisation with behaviour in autism: a multimodal EU-AIMS Longitudinal European Autism Project (LEAP) study". In: *Molecular Autism* 14.1 (Aug. 2023), p. 32. issn: 2040-2392. doi: 10.1186/s13229-023-00564-3. url: https://doi.org/10.1186/s13229-023-00564-3 (visited on 04/11/2024).

[272] Kyoung-Su Oh and Keechul Jung. "GPU implementation of neural networks". In: *Pattern Recognition* 37.6 (June 2004), pp. 1311–1314. issn: 0031-3203. doi: 10.1016/j.patcog.2004.01.013. url: https://www.sciencedirect.com/science/article/pii/S0031320304000524 (visited on 02/29/2024).

[273] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. *Representation Learning with Contrastive Predictive Coding*. arXiv:1807.03748 [cs, stat]. Jan. 2019. doi: 10.48550/arXiv.1807.03748. url: http://arxiv.org/abs/1807.03748 (visited on 02/20/2024).

[274] Emanuele Palumbo, Imant Daunhawer, and Julia E. Vogt. "MMVAE+: Enhancing the Generative Quality of Multimodal VAEs without Compromises". en. In: *The Eleventh International Conference on Learning Representations, {ICLR} 2023, Kigali, Rwanda, May 1-5, 2023*. Mar. 2022. url: https://openreview.net/forum?id=B42UJTVdDZ5 (visited on 08/03/2023).

[275] Gaurav Pandey and Ambedkar Dukkipati. "Variational methods for conditional multimodal deep learning". en. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. Anchorage, AK, USA: IEEE, May 2017, pp. 308–315. isbn: 978-1-5090-6182-2. doi: 10.1109/IJCNN.2017.7965870. url: http://ieeexplore.ieee.org/document/7965870/ (visited on 03/19/2024).

[276] Linden Parkes, Tyler M. Moore, Monica E. Calkins, Philip A. Cook, Matthew Cieslak, David R. Roalf, Daniel H. Wolf, Ruben C. Gur, Raquel E. Gur, Theodore D. Satterthwaite, and Danielle S. Bassett. "Transdiagnostic dimensions of psychopathology explain individuals' unique deviations from normative neurodevelopment in brain structure".

en. In: *Translational Psychiatry* 11.1 (Apr. 2021). Publisher: Nature Publishing Group, pp. 1–13. issn: 2158-3188. doi: 10.1038/s41398-021-01342-6. url: https://www.nature.com/articles/s41398-021-01342-6 (visited on 04/11/2024).

[277] Prasanna Parvathaneni, Shunxing Bao, Vishwesh Nath, Neil D. Woodward, Daniel O. Claassen, Carissa J. Cascio, David H. Zald, Yuankai Huo, Bennett A. Landman, and Ilwoo Lyu. "Cortical Surface Parcellation Using Spherical Convolutional Neural Networks". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 501–509. isbn: 978-3-030-32248-9. doi: 10.1007/978-3-030-32248-9_56.

[278] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. "Context Encoders: Feature Learning by Inpainting". English. In: *2016 {IEEE} Conference on Computer Vision and Pattern Recognition, {CVPR} 2016, Las Vegas, NV, USA, June 27-30, 2016*. ISSN: 1063-6919. IEEE Computer Society, June 2016, pp. 2536–2544. isbn: 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.278. url: https://www.computer.org/csdl/proceedings-article/cvpr/2016/8851c536/12OmNCgJe72 (visited on 03/06/2024).

[279] Nancy Pedano, Adam E. Flanders, Lisa Scarpace, Tom Mikkelsen, Jennifer M. Eschbacher, Beth Hermes, Victor Sisneros, Jill Barnholtz-Sloan, and Quinn Ostrom. *The Cancer Genome Atlas Low Grade Glioma Collection (TCGA-LGG)*. 2016. doi: 10.7937/K9/TCIA.2016.L4LTD3TK. url: https://www.cancerimagingarchive.net/collection/tcga-lgg/ (visited on 04/08/2024).

[280] Helena Pelin, Marcus Ising, Frederike Stein, Susanne Meinert, Tina Meller, Katharina Brosch, Nils R. Winter, Axel Krug, Ramona Leenings, Hannah Lemke, Igor Nenadić, Stefanie Heilmann-Heimbach, Andreas J. Forstner, Markus M. Nöthen, Nils Opel, Jonathan Repple, Julia Pfarr, Kai Ringwald, Simon Schmitt, Katharina Thiel, Lena Waltemate, Alexandra Winter, Fabian Streit, Stephanie Witt, Marcella Rietschel, Udo Dannlowski, Tilo Kircher, Tim Hahn, Bertram Müller-Myhsok, and Till F. M. Andlauer. "Identification of transdiagnostic psychiatric disorder subtypes using unsupervised learning". en. In: *Neuropsychopharmacology* 46.11 (Oct. 2021). Publisher: Nature Publishing Group, pp. 1895–1905. issn: 1740-634X. doi: 10.1038/s41386-021-01051-0. url: https://www.nature.com/articles/s41386-021-01051-0 (visited on 03/12/2024).

[281] Matthieu Perrot, Denis Rivière, and Jean-François Mangin. "Cortical sulci recognition and spatial normalization". In: *Medical Image Analysis*. Special section on IPMI 2009 15.4 (Aug. 2011), pp. 529–550. issn: 1361-8415. doi: 10.1016/j.media.2011.02.008. url: https://www.sciencedirect.com/science/article/pii/S1361841511000302 (visited on 04/07/2024).

[282] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. "Alzheimer's Disease Neuroimaging Initiative (ADNI)". In: *Neurology* 74.3 (Jan. 2010).

Publisher: Wolters Kluwer, pp. 201–209. doi: 10.1212/WNL.0b013e3181cb3e25. url: https://www.neurology.org/doi/10.1212/WNL.0b013e3181cb3e25 (visited on 04/08/2024).

[283] Clement Poiret, Antoine Bouyeure, Sandesh Patil, Cécile Boniteau, Edouard Duchesnay, Antoine Grigis, Frederic Lemaitre, and Marion Noulhiane. "Attention-gated 3D CapsNet for robust hippocampal segmentation". In: *Journal of Medical Imaging* 11.1 (Jan. 2024). Publisher: SPIE, p. 014003. issn: 2329-4302, 2329-4310. doi: 10.1117/1.JMI.11.1.014003. url: https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-11/issue-1/014003/Attention-gated-3D-CapsNet-for-robust-hippocampal-segmentation/10.1117/1.JMI.11.1.014003.full (visited on 04/18/2024).

[284] Richie Poulton, Terrie E. Moffitt, and Phil A. Silva. "The Dunedin Multidisciplinary Health and Development Study: Overview of the first 40 years, with an eye to the future". In: *Social Psychiatry and Psychiatric Epidemiology: The International Journal for Research in Social and Genetic Epidemiology and Mental Health Services* 50.5 (2015). Place: Germany Publisher: Springer, pp. 679–693. issn: 1433-9285. doi: 10.1007/s00127-015-1048-8.

[285] Lin Qiu, Lynn Lin, and Vernon M. Chinchilli. *Variational Interpretable Learning from Multi-view Data*. arXiv:2202.13503 [cs, stat]. Mar. 2022. doi: 10.48550/arXiv.2202.13503. url: http://arxiv.org/abs/2202.13503 (visited on 09/29/2023).

[286] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. "Transfusion: Understanding Transfer Learning for Medical Imaging". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019. url: https://proceedings.neurips.cc/paper_files/paper/2019/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html (visited on 01/18/2024).

[287] Vidya Rajagopalan, Julia Scott, Piotr A. Habas, Kio Kim, James Corbett-Detig, Francois Rousseau, A. James Barkovich, Orit A. Glenn, and Colin Studholme. "Local Tissue Growth Patterns Underlying Normal Fetal Human Brain Gyrification Quantified In Utero". en. In: *Journal of Neuroscience* 31.8 (Feb. 2011). Publisher: Society for Neuroscience Section: Articles, pp. 2878–2887. issn: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.5458-10.2011. url: https://www.jneurosci.org/content/31/8/2878 (visited on 04/03/2024).

[288] Pasko Rakic. "Evolution of the neocortex: a perspective from developmental biology". en. In: *Nature Reviews Neuroscience* 10.10 (Oct. 2009). Publisher: Nature Publishing Group, pp. 724–735. issn: 1471-0048. doi: 10.1038/nrn2719. url: https://www.nature.com/articles/nrn2719 (visited on 04/03/2024).

[289] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. "Zero-Shot Text-to-Image Generation". en. In: *Proceedings of the 38th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR,

July 2021, pp. 8821–8831. url: https://proceedings.mlr.press/v139/ramesh21a.html (visited on 03/07/2024).

[290] Amine Rebei, Agusti Alentorn, Hamza Chegraoui, Vincent Frouin, and Cathy Philippe. "Contribution Of Imaging-Genetics To Overall Survival Prediction Compared To Clinical Status For Pcnsl Patients". In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. ISSN: 1945-8452. Apr. 2021, pp. 832–835. doi: 10.1109/ISBI48211.2021.9433785. url: https://ieeexplore.ieee.org/document/9433785 (visited on 03/26/2024).

[291] Fernanda L. Ribeiro, Steffen Bollmann, Ross Cunnington, and Alexander M. Puckett. *An explainability framework for cortical surface-based deep learning*. arXiv:2203.08312 [cs, q-bio]. Mar. 2022. url: http://arxiv.org/abs/2203.08312 (visited on 04/03/2024).

[292] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". en. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM, Aug. 2016, pp. 1135–1144. isbn: 978-1-4503-4232-2. doi: 10.1145/2939672.2939778. url: https://dl.acm.org/doi/10.1145/2939672.2939778 (visited on 03/06/2024).

[293] Herbert Robbins and Sutton Monro. "A Stochastic Approximation Method". In: *The Annals of Mathematical Statistics* 22.3 (Sept. 1951). Publisher: Institute of Mathematical Statistics, pp. 400–407. issn: 0003-4851, 2168-8990. doi: 10.1214/aoms/1177729586. url: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-22/issue-3/A-Stochastic-Approximation-Method/10.1214/aoms/1177729586.full (visited on 02/29/2024).

[294] Nikolaos Rodis, Christos Sardianos, Georgios Th Papadopoulos, Panagiotis Radoglou-Grammatikis, Panagiotis Sarigiannidis, and Iraklis Varlamis. *Multimodal Explainable Artificial Intelligence: A Comprehensive Review of Methodological Advances and Future Research Directions*. arXiv:2306.05731 [cs]. June 2023. doi: 10.48550/arXiv.2306.05731. url: http://arxiv.org/abs/2306.05731 (visited on 03/15/2024).

[295] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-Resolution Image Synthesis With Latent Diffusion Models". en. In: *{IEEE/CVF} Conference on Computer Vision and Pattern Recognition, {CVPR} 2022, New Orleans, LA, USA, June 18-24, 2022*. 2022, pp. 10684–10695. url: https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html (visited on 03/01/2024).

[296] A. L. Romer, A. R. Knodt, R. Houts, B. D. Brigidi, T. E. Moffitt, A. Caspi, and A. R. Hariri. "Structural alterations within cerebellar circuitry are associated with general liability for common mental disorders". en. In: *Molecular Psychiatry* 23.4 (Apr. 2018). Number: 4 Publisher: Nature Publishing Group, pp. 1084–1090. issn: 1476-5578. doi: 10.1038/mp.2017.57. url: https://www.nature.com/articles/mp201757 (visited on 02/01/2024).

[297] Adrienne L. Romer, Annchen R. Knodt, Maria L. Sison, David Ireland, Renate Houts, Sandhya Ramrakha, Richie Poulton, Ross Keenan, Tracy R. Melzer, Terrie E. Moffitt, Avshalom Caspi, and Ahmad R. Hariri. "Replicability of structural brain alterations associated with general psychopathology: evidence from a population-representative birth cohort". en. In: *Molecular Psychiatry* 26.8 (Aug. 2021). Number: 8 Publisher: Nature Publishing Group, pp. 3839–3846. issn: 1476-5578. doi: 10.1038/s41380-019-0621-z. url: https://www.nature.com/articles/s41380-019-0621-z (visited on 02/01/2024).

[298] Adrienne L. Romer and Diego A. Pizzagalli. "Associations Between Brain Structural Alterations, Executive Dysfunction, and General Psychopathology in a Healthy and Cross-Diagnostic Adult Patient Sample". In: *Biological Psychiatry Global Open Science* 2.1 (Jan. 2022), pp. 17–27. issn: 2667-1743. doi: 10.1016/j.bpsgos.2021.06.002. url: https://www.sciencedirect.com/science/article/pii/S2667174321000483 (visited on 04/02/2024).

[299] Nanda Rommelse, Jan K. Buitelaar, and Catharina A. Hartman. "Structural brain imaging correlates of ASD and ADHD across the lifespan: a hypothesis-generating review on developmental ASD–ADHD subtypes". en. In: *Journal of Neural Transmission* 124.2 (Feb. 2017), pp. 259–271. issn: 1435-1463. doi: 10.1007/s00702-016-1651-1. url: https://doi.org/10.1007/s00702-016-1651-1 (visited on 04/11/2024).

[300] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". en. In: *arXiv:1505.04597 [cs]* (May 2015). arXiv: 1505.04597. url: http://arxiv.org/abs/1505.04597 (visited on 11/30/2020).

[301] Adon F. G. Rosen, David R. Roalf, Kosha Ruparel, Jason Blake, Kevin Seelaus, Lakshmi P. Villa, Rastko Ciric, Philip A. Cook, Christos Davatzikos, Mark A. Elliott, Angel Garcia de La Garza, Efstathios D. Gennatas, Megan Quarmley, J. Eric Schmitt, Russell T. Shinohara, M. Dylan Tisdall, R. Cameron Craddock, Raquel E. Gur, Ruben C. Gur, and Theodore D. Satterthwaite. "Quantitative assessment of structural image quality". In: *NeuroImage* 169 (Apr. 2018), pp. 407–418. issn: 1053-8119. doi: 10.1016/j.neuroimage.2017.12.059. url: https://www.sciencedirect.com/science/article/pii/S1053811917310832 (visited on 04/03/2024).

[302] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain". In: *Psychological Review* 65.6 (1958). Place: US Publisher: American Psychological Association, pp. 386–408. issn: 1939-1471. doi: 10.1037/h0042519.

[303] Roman Rosipal and Nicole Krämer. "Overview and Recent Advances in Partial Least Squares". en. In: *Subspace, Latent Structure and Feature Selection*. Ed. by Craig Saunders, Marko Grobelnik, Steve Gunn, and John Shawe-Taylor. Berlin, Heidelberg: Springer, 2006, pp. 34–51. isbn: 978-3-540-34138-3. doi: 10.1007/11752790_2.

[304] John L.R. Rubenstein and Pasko Rakic. "Genetic Control of Cortical Development". In: *Cerebral Cortex* 9.6 (Sept. 1999), pp. 521–523. issn: 1047-3211. doi: 10.1093/cercor/9.6.521. url: https://doi.org/10.1093/cercor/9.6.521 (visited on 03/28/2024).

[305] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". en. In: *Nature* 323.6088 (Oct. 1986). Number: 6088 Publisher: Nature Publishing Group, pp. 533–536. issn: 1476-4687. doi: 10.1038/323533a0. url: https://www.nature.com/articles/323533a0 (visited on 03/04/2024).

[306] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. "Dynamic Routing Between Capsules". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. url: https://proceedings.neurips.cc/paper_files/paper/2017/hash/2cad8fa47bbef282badbb8de5374b894-Abstract.html (visited on 03/01/2024).

[307] Nikunj Saunshi, Jordan Ash, Surbhi Goel, Dipendra Misra, Cyril Zhang, Sanjeev Arora, Sham Kakade, and Akshay Krishnamurthy. "Understanding Contrastive Learning Requires Incorporating Inductive Biases". en. In: *Proceedings of the 39th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, June 2022, pp. 19250–19286. url: https://proceedings.mlr.press/v162/saunshi22a.html (visited on 01/26/2024).

[308] Lisa Scarpace, Tom Mikkelsen, Soonmee Cha, Sujaya Rao, Sangeeta Tekchandani, David Gutman, Joel H. Saltz, Bradley J. Erickson, Nancy Pedano, Adam E. Flanders, Jill Barnholtz-Sloan, Quinn Ostrom, Daniel Barboriak, and Laura J. Pierce. *The Cancer Genome Atlas Glioblastoma Multiforme Collection (TCGA-GBM)*. 2016. doi: 10.7937/K9/TCIA.2016.RNYFUYE9. url: https://www.cancerimagingarchive.net/collection/tcga-gbm/ (visited on 04/08/2024).

[309] Carmen Schaeuffele, Laura E. Meine, Ava Schulz, Maxi C. Weber, Angela Moser, Christina Paersch, Dominique Recher, Johanna Boettcher, Babette Renneberg, Christoph Flückiger, and Birgit Kleim. "A systematic review and meta-analysis of transdiagnostic cognitive behavioural therapies for emotional disorders". en. In: *Nature Human Behaviour* 8.3 (Mar. 2024). Publisher: Nature Publishing Group, pp. 493–509. issn: 2397-3374. doi: 10.1038/s41562-023-01787-3. url: https://www.nature.com/articles/s41562-023-01787-3 (visited on 04/11/2024).

[310] Gunter Schumann, Elisabeth B. Binder, et al. "Stratified medicine for mental disorders". In: *European Neuropsychopharmacology* 24.1 (Jan. 2014), pp. 5–50. issn: 0924-977X. doi: 10.1016/j.euroneuro.2013.09.010. url: https://www.sciencedirect.com/science/article/pii/S0924977X13002769 (visited on 03/12/2024).

[311] Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis Vlahavas. "On the Stratification of Multi-label Data". en. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis. Berlin, Heidelberg: Springer, 2011, pp. 145–158. isbn: 978-3-642-23808-6. doi: 10.1007/978-3-642-23808-6_10.

[312] Gabriela Sejnova, Michal Vavrecka, and Karla Stepanova. *Benchmarking Multimodal Variational Autoencoders: CdSprites+ Dataset and Toolkit*. arXiv:2209.03048 [cs]. Nov. 2023. doi: 10.48550/arXiv.2209.03048. url: http://arxiv.org/abs/2209.03048 (visited on 03/08/2024).

[313] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. ISSN: 2380-7504. Oct. 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74. url: https://ieeexplore.ieee.org/document/8237336 (visited on 03/06/2024).

[314] Si-Baek Seong, Chongwon Pae, and Hae-Jeong Park. "Geometric Convolutional Neural Network for Analyzing Surface-Based Neuroimaging Data". In: *Frontiers in Neuroinformatics* 12 (2018), p. 42. issn: 1662-5196. doi: 10.3389/fninf.2018.00042. url: https://www.frontiersin.org/article/10.3389/fninf.2018.00042 (visited on 11/10/2021).

[315] Harshay Shah, Kaustav Tamuly, Aditi Raghunathan, Prateek Jain, and Praneeth Netrapalli. "The Pitfalls of Simplicity Bias in Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 9573–9585. url: https://proceedings.neurips.cc/paper/2020/hash/6cfe0e6127fa25df2a0ef2ae1067d915-Abstract.html (visited on 03/07/2024).

[316] Julia M. Sheffield, Sridhar Kandala, Carol A. Tamminga, Godfrey D. Pearlson, Matcheri S. Keshavan, John A. Sweeney, Brett A. Clementz, Dov B. Lerman-Sinkoff, S. Kristian Hill, and Deanna M. Barch. "Transdiagnostic Associations Between Functional Brain Network Integrity and Cognition". In: *JAMA Psychiatry* 74.6 (June 2017), pp. 605–613. issn: 2168-622X. doi: 10.1001/jamapsychiatry.2017.0669. url: https://doi.org/10.1001/jamapsychiatry.2017.0669 (visited on 04/11/2024).

[317] Yuge Shi, Brooks Paige, Philip Torr, and Siddharth N. "Relating by Contrasting: A Data-efficient Framework for Multimodal Generative Models". en. In: *9th International Conference on Learning Representations, {ICLR} 2021, Virtual Event, Austria, May 3-7, 2021*. Oct. 2020. url: https://openreview.net/forum?id=vhKe9UFbrJo (visited on 07/12/2023).

[318] Yuge Shi, N. Siddharth, Brooks Paige, and Philip H. S. Torr. "Variational Mixture-of-Experts Autoencoders for Multi-Modal Deep Generative Models". In: *arXiv:1911.03393 [cs, stat]* (Nov. 2019). arXiv: 1911.03393. url: http://arxiv.org/abs/1911.03393 (visited on 03/16/2021).

[319] Ravid Shwartz-Ziv and Yann LeCun. *To Compress or Not to Compress- Self-Supervised Learning and Information Theory: A Review*. arXiv:2304.09355 [cs, math]. Nov. 2023. url: http://arxiv.org/abs/2304.09355 (visited on 03/06/2024).

[320] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps". en. In: *2nd International Conference on Learning Representations, {ICLR} 2014, Banff, AB, Canada, April 14-16, 2014, Workshop Track Proceedings*. Dec. 2013. url: https://openreview.net/forum?id=c04ycnpqxKcS9 (visited on 03/06/2024).

[321] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *Conference Track Proceedings*. arXiv:1409.1556 [cs]. San Diego, CA, USA: ICLR, May 2015. doi: 10.48550/arXiv.1409.1556. url: http://arxiv.org/abs/1409.1556 (visited on 01/30/2024).

[322] Marion Smits. "MRI biomarkers in neuro-oncology". en. In: *Nature Reviews Neurology* 17.8 (Aug. 2021). Publisher: Nature Publishing Group, pp. 486–500. issn: 1759-4766. doi: 10.1038/s41582-021-00510-y. url: https://www.nature.com/articles/s41582-021-00510-y (visited on 04/07/2024).

[323] Kihyuk Sohn. "Improved Deep Metric Learning with Multi-class N-pair Loss Objective". In: *Advances in Neural Information Processing Systems*. Vol. 29. Curran Associates, Inc., 2016. url: https://papers.nips.cc/paper_files/paper/2016/hash/6b180037abbebea991d8b1232f8a8ca9-Abstract.html (visited on 02/20/2024).

[324] Marco Solmi, Joaquim Radua, Miriam Olivola, Enrico Croce, Livia Soardo, Gonzalo Salazar de Pablo, Jae Il Shin, James B. Kirkbride, Peter Jones, Jae Han Kim, Jong Yeob Kim, Andrè F. Carvalho, Mary V. Seeman, Christoph U. Correll, and Paolo Fusar-Poli. "Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies". en. In: *Molecular Psychiatry* 27.1 (Jan. 2022). Publisher: Nature Publishing Group, pp. 281–295. issn: 1476-5578. doi: 10.1038/s41380-021-01161-7. url: https://www.nature.com/articles/s41380-021-01161-7 (visited on 03/13/2024).

[325] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. "Deep Metric Learning via Lifted Structured Feature Embedding". en. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, June 2016, pp. 4004–4012. isbn: 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.434. url: http://ieeexplore.ieee.org/document/7780803/ (visited on 02/20/2024).

[326] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. issn: 1533-7928. url: http://jmlr.org/papers/v15/srivastava14a.html (visited on 03/20/2024).

[327] Nitish Srivastava and Ruslan Salakhutdinov. "Learning Representations for Multimodal Data with Deep Belief Nets". en. In: *ICML Representation Learning Workshop, Edinburgh, Scotland, UK, 2012*. 2012.

[328] Nitish Srivastava and Russ R. Salakhutdinov. "Multimodal Learning with Deep Boltzmann Machines". en. In: *Advances in Neural Information Processing Systems* 25 (2012). url: https://proceedings.neurips.cc/paper/2012/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html (visited on 03/25/2021).

[329] Sören Richard Stahlschmidt, Benjamin Ulfenborg, and Jane Synnergren. "Multimodal deep learning for biomedical data fusion: a review". In: *Briefings in Bioinformatics* 23.2 (Mar. 2022), bbab569. issn: 1477-4054. doi: 10.1093/bib/bbab569. url: https://doi.org/10.1093/bib/bbab569 (visited on 03/26/2024).

[330] Laura Stefanik, Lauren Erdman, Stephanie H. Ameis, George Foussias, Benoit H. Mulsant, Tina Behdinan, Anna Goldenberg, Lauren J. O'Donnell, and Aristotle N. Voineskos. "Brain-Behavior Participant Similarity Networks Among Youth and Emerging Adults with Schizophrenia Spectrum, Autism Spectrum, or Bipolar Disorder and Matched Controls". en. In: *Neuropsychopharmacology* 43.5 (Apr. 2018). Publisher: Nature Publishing Group, pp. 1180–1188. issn: 1740-634X. doi: 10.1038/npp.2017.274. url: https://www.nature.com/articles/npp2017274 (visited on 04/11/2024).

[331] Jan Stochl, Hannah Jones, Emma Soneson, Adam P. Wagner, Golam M. Khandaker, Stanley Zammit, Jon Heron, Gemma Hammerton, Edward T. Bullmore, Ray Dolan, Peter Fonagy, Ian M. Goodyer, J. Perez, and Peter B. Jones. "Stratification of adolescents across mental phenomena emphasizes the importance of transdiagnostic distress: a replication in two general population cohorts". en. In: *European Child & Adolescent Psychiatry* 32.5 (May 2023), pp. 797–807. issn: 1435-165X. doi: 10.1007/s00787-021-01909-0. url: https://doi.org/10.1007/s00787-021-01909-0 (visited on 03/12/2024).

[332] Georg F. Striedter, Shyam Srinivasan, and Edwin S. Monuki. "Cortical Folding: When, Where, How, and Why?" fr. In: *Annual Review of Neuroscience* 38.Volume 38, 2015 (July 2015). Publisher: Annual Reviews, pp. 291–307. issn: 0147-006X, 1545-4126. doi: 10.1146/annurev-neuro-071714-034128. url: https://www.annualreviews.org/content/journals/10.1146/annurev-neuro-071714-034128 (visited on 04/03/2024).

[333] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis". In: *NeuroImage* 101 (Nov. 2014), pp. 569–582. issn: 1053-8119. doi: 10.1016/j.neuroimage.2014.06.077. url: https://www.sciencedirect.com/science/article/pii/S1053811914005540 (visited on 03/27/2024).

[334] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to Sequence Learning with Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc., 2014. url: https://papers.nips.cc/paper_files/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html (visited on 03/04/2024).

[335] Thomas Sutter, Imant Daunhawer, and Julia Vogt. "Multimodal Generative Learning Utilizing Jensen-Shannon-Divergence". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 6100–6110. url: https://proceedings.neurips.cc/paper/2020/hash/43bb733c1b62a5e374c63cb22fa457b4-Abstract.html (visited on 03/19/2024).

[336] Thomas M. Sutter, Imant Daunhawer, and Julia E. Vogt. "Generalized Multimodal ELBO". en. In: *9th International Conference on Learning Representations, {ICLR} 2021, Virtual Event, Austria, May 3-7, 2021*. Oct. 2020. url: https://openreview.net/forum?id=5Y21V0RDBV (visited on 08/03/2023).

[337] Masahiro Suzuki and Yutaka Matsuo. "A survey of multimodal deep generative models". In: *Advanced Robotics* 36.5-6 (Mar. 2022). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01691864.2022.2035253, pp. 261–278. issn: 0169-1864. doi: 10.1080/01691864.2022.2035253. url: https://doi.org/10.1080/01691864.2022.2035253 (visited on 03/19/2024).

[338] Masahiro Suzuki, Kotaro Nakayama, and Yutaka Matsuo. "Joint Multimodal Learning with Deep Generative Models". In: *arXiv:1611.01891 [cs, stat]* (Nov. 2016). arXiv: 1611.01891. url: http://arxiv.org/abs/1611.01891 (visited on 03/16/2021).

[339] Valerie J. Sydnor, Bart Larsen, Danielle S. Bassett, Aaron Alexander-Bloch, Damien A. Fair, Conor Liston, Allyson P. Mackey, Michael P. Milham, Adam Pines, David R. Roalf, Jakob Seidlitz, Ting Xu, Armin Raznahan, and Theodore D. Satterthwaite. "Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology". In: *Neuron* 109.18 (Sept. 2021), pp. 2820–2846. issn: 0896-6273. doi: 10.1016/j.neuron.2021.06.016. url: https://www.sciencedirect.com/science/article/pii/S0896627321004578 (visited on 03/28/2024).

[340] Carol A. Tamminga, Godfrey Pearlson, Matcheri Keshavan, John Sweeney, Brett Clementz, and Gunvant Thaker. "Bipolar and Schizophrenia Network for Intermediate Phenotypes: Outcomes Across the Psychosis Continuum". In: *Schizophrenia Bulletin* 40.Suppl_2 (Mar. 2014), S131–S137. issn: 0586-7614. doi: 10.1093/schbul/sbt179. url: https://doi.org/10.1093/schbul/sbt179 (visited on 03/13/2024).

[341] Gemini Team, Rohan Anil, et al. *Gemini: A Family of Highly Capable Multimodal Models*. arXiv:2312.11805 [cs]. Dec. 2023. doi: 10.48550/arXiv.2312.11805. url: http://arxiv.org/abs/2312.11805 (visited on 04/02/2024).

[342] Arthur Tenenhaus, Cathy Philippe, Vincent Guillemot, Kim-Anh Le Cao, Jacques Grill, and Vincent Frouin. "Variable selection for generalized canonical correlation analysis". In: *Biostatistics* 15.3 (July 2014), pp. 569–583. issn: 1465-4644. doi: 10.1093/biostatistics/kxu001. url: https://doi.org/10.1093/biostatistics/kxu001 (visited on 03/14/2024).

[343] Arthur Tenenhaus and Michel Tenenhaus. "Regularized Generalized Canonical Correlation Analysis". en. In: *Psychometrika* 76.2 (Apr. 2011), pp. 257–284. issn: 1860-0980. doi: 10.1007/s11336-011-9206-8. url: https://doi.org/10.1007/s11336-011-9206-8 (visited on 03/19/2024).

[344] M. Tenenhaus. "L'approche PLS". fr. In: *Revue de Statistique Appliquée* 47.2 (1999), pp. 5–40. issn: 2400-4812. url: http://archive.numdam.org/item/RSA_1999__47_2_5_0/ (visited on 03/14/2024).

[345] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. "What Makes for Good Views for Contrastive Learning?" In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 6827–6839. url: https://proceedings.neurips.cc/paper/2020/hash/4c2e5eaae9152079b9e95845750bb9ab-Abstract.html (visited on 01/26/2024).

[346] Robert Tibshirani. "Regression Shrinkage and Selection Via the Lasso". en. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1996.tb02080.x, pp. 267–288. issn: 2517-6161. doi: 10.1111/j.2517-6161.1996.tb02080.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x (visited on 03/21/2024).

[347] Xiaoyu Tong, Hua Xie, Nancy Carlisle, Gregory A. Fonzo, Desmond J. Oathes, Jing Jiang, and Yu Zhang. "Transdiagnostic connectome signatures from resting-state fMRI predict individual-level intellectual capacity". en. In: *Translational Psychiatry* 12.1 (Sept. 2022). Publisher: Nature Publishing Group, pp. 1–11. issn: 2158-3188. doi: 10.1038/s41398-022-02134-2. url: https://www.nature.com/articles/s41398-022-02134-2 (visited on 04/11/2024).

[348] Yao-Hung Hubert Tsai, Paul Pu Liang, Amir Zadeh, Louis-Philippe Morency, and Ruslan Salakhutdinov. "Learning Factorized Multimodal Representations". en. In: *7th International Conference on Learning Representations, {ICLR} 2019, New Orleans, LA, USA, May 6-9, 2019*. Sept. 2018. url: https://openreview.net/forum?id=rygqqsA9KX (visited on 03/20/2024).

[349] Guillermo Valle-Perez, Chico Q. Camargo, and Ard A. Louis. "Deep learning generalizes because the parameter-function map is biased towards simple functions". en. In: *7th International Conference on Learning Representations, {ICLR} 2019, New Orleans, LA, USA, May 6-9, 2019*. Sept. 2018. url: https://openreview.net/forum?id=rye4g3AqFm (visited on 03/08/2024).

[350] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. url: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (visited on 01/31/2024).

[351] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. "Graph Attention Networks". en. In: *6th International Conference on Learning Representations, {ICLR} 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. Feb. 2018. url: https://openreview.net/forum?id=rJXMpikCZ (visited on 01/31/2024).

[352] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and composing robust features with denoising autoencoders". In: *Proceedings of the 25th international conference on Machine learning*. ICML '08. New York, NY, USA: Association for Computing Machinery, July 2008, pp. 1096–1103. isbn: 978-1-60558-205-4. doi: 10.1145/1390156.1390294. url: https://doi.org/10.1145/1390156.1390294 (visited on 03/06/2024).

[353] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion". In: *Journal of Machine Learning Research* 11.110 (2010), pp. 3371–3408. issn: 1533-7928. url: http://jmlr.org/papers/v11/vincent10a.html (visited on 03/06/2024).

[354] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and tell: A neural image caption generator". en. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, June 2015, pp. 3156–3164. isbn: 978-1-4673-6964-0. doi: 10.1109/CVPR.2015.7298935. url: http://ieeexplore.ieee.org/document/7298935/ (visited on 03/04/2024).

[355] Brent A. Vogt, David M. Finch, and Carl R. Olson. "Functional Heterogeneity in Cingulate Cortex: The Anterior Executive and Posterior Evaluative Regions". In: *Cerebral Cortex* 2.6 (Nov. 1992), pp. 435–443. issn: 1047-3211. doi: 10.1093/cercor/2.6.435-a. url: https://doi.org/10.1093/cercor/2.6.435-a (visited on 02/09/2024).

[356] Maria Vounou, Thomas E. Nichols, and Giovanni Montana. "Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach". In: *NeuroImage*. Imaging Genetics 53.3 (Nov. 2010), pp. 1147–1159. issn: 1053-8119. doi: 10.1016/j.neuroimage.2010.07.002. url: https://www.sciencedirect.com/science/article/pii/S1053811910009535 (visited on 03/14/2024).

[357] Christian Wachinger, Anna Rieckmann, and Sebastian Pölsterl. "Detect and correct bias in multi-site neuroimaging datasets". In: *Medical Image Analysis* 67 (Jan. 2021), p. 101879. issn: 1361-8415. doi: 10.1016/j.media.2020.101879. url: https://www.sciencedirect.com/science/article/pii/S1361841520302437 (visited on 02/19/2024).

[358] Chong Wang. "Variational Bayesian Approach to Canonical Correlation Analysis". In: *IEEE Transactions on Neural Networks* 18.3 (2007). Conference Name: IEEE Transactions on Neural Networks, pp. 905–910. issn: 1941-0093. doi: 10.1109/TNN.2007.891186. url: https://ieeexplore.ieee.org/document/4182407 (visited on 03/14/2024).

[359] Feng Wang and Huaping Liu. "Understanding the Behaviour of Contrastive Loss". en. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE, June 2021, pp. 2495–2504. isbn: 978-1-66544-509-2. doi: 10.1109/CVPR46437.2021.00252. url: https://ieeexplore.ieee.org/document/9577669/ (visited on 01/26/2024).

[360] Lei Wang, Kathryn I. Alpert, Vince D. Calhoun, Derin J. Cobia, David B. Keator, Margaret D. King, Alexandr Kogan, Drew Landis, Marcelo Tallis, Matthew D. Turner, Steven G. Potkin, Jessica A. Turner, and Jose Luis Ambite. "SchizConnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration". In: *NeuroImage*. Sharing the wealth: Brain Imaging Repositories in 2015 124 (Jan. 2016), pp. 1155–1167. issn: 1053-8119. doi: 10.1016/j.neuroimage.2015.06.065. url: https:

//www.sciencedirect.com/science/article/pii/S1053811915005704 (visited on 03/13/2024).

[361] Weiran Wang, Raman Arora, Karen Livescu, and Jeff Bilmes. "On deep multi-view representation learning". In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. ICML'15. Lille, France: JMLR.org, July 2015, pp. 1083–1092. (Visited on 03/19/2024).

[362] Weiran Wang, Xinchen Yan, Honglak Lee, and Karen Livescu. "Deep Variational Canonical Correlation Analysis". en. In: *CoRR* (Nov. 2016). url: https://openreview.net/forum?id=H1Heentlx (visited on 03/19/2024).

[363] Xuetong Wang, Kanhao Zhao, Rong Zhou, Alex Leow, Ricardo Osorio, Yu Zhang, and Lifang He. *Normative Modeling via Conditional Variational Autoencoder and Adversarial Learning to Identify Brain Dysfunction in Alzheimer's Disease*. arXiv:2211.08982 [cs, q-bio]. Nov. 2022. doi: 10.48550/arXiv.2211.08982. url: http://arxiv.org/abs/2211.08982 (visited on 04/18/2024).

[364] Ethan Weinberger, Nicasia Beebe-Wang, and Su-In Lee. "Moment Matching Deep Contrastive Latent Variable Models". en. In: *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*. ISSN: 2640-3498. PMLR, May 2022, pp. 2354–2371. url: https://proceedings.mlr.press/v151/weinberger22a.html (visited on 02/20/2024).

[365] John N. Weinstein, Eric A. Collisson, Gordon B. Mills, Kenna R. Mills Shaw, Brad A. Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M. Stuart. "The Cancer Genome Atlas Pan-Cancer analysis project". en. In: *Nature Genetics* 45.10 (Oct. 2013). Publisher: Nature Publishing Group, pp. 1113–1120. issn: 1546-1718. doi: 10.1038/ng.2764. url: https://www.nature.com/articles/ng.2764 (visited on 03/27/2024).

[366] P. J. Werbos. "Applications of advances in nonlinear sensitivity analysis". English. In: *Proceedings of the 10th IFIP Conference, 31.8 - 4.9, NYC*. 1981. url: https://www.osti.gov/etdeweb/biblio/5080493 (visited on 02/29/2024).

[367] Udaranga Wickramasinghe, Edoardo Remelli, Graham Knott, and Pascal Fua. "Voxel2Mesh: 3D Mesh Model Generation from Volumetric Data". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz. Cham: Springer International Publishing, 2020, pp. 299–308. isbn: 978-3-030-59719-1. doi: 10.1007/978-3-030-59719-1_30.

[368] Leanne M. Williams. "Special Report: Precision Psychiatry—Are We Getting Closer?" en. In: *Psychiatric News* (Aug. 2022). Publisher: American Psychiatric Publishing, Inc. doi: 10.1176/appi.pn.2022.09.9.23. url: https://psychnews.psychiatryonline.org/doi/10.1176/appi.pn.2022.09.9.23 (visited on 04/11/2024).

[369] Mitchell Wilson. "DSM-III and the transformation of American psychiatry: a history". In: *American Journal of Psychiatry* 150.3 (Mar. 1993). Publisher: American Psychiatric Publishing, pp. 399–410. issn: 0002-953X. doi: 10.1176/ajp.150.3.399. url: https://ajp.psychiatryonline.org/doi/10.1176/ajp.150.3.399 (visited on 03/12/2024).

[370] Daniela M. Witten, Robert Tibshirani, and Trevor Hastie. "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis". In: *Biostatistics* 10.3 (July 2009), pp. 515–534. issn: 1465-4644. doi: 10.1093/biostatistics/kxp008. url: https://doi.org/10.1093/biostatistics/kxp008 (visited on 03/14/2024).

[371] Svante Wold, Michael Sjöström, and Lennart Eriksson. "PLS-regression: a basic tool of chemometrics". In: *Chemometrics and Intelligent Laboratory Systems*. PLS Methods 58.2 (Oct. 2001), pp. 109–130. issn: 0169-7439. doi: 10.1016/S0169-7439(01)00155-1. url: https://www.sciencedirect.com/science/article/pii/S0169743901001551 (visited on 03/13/2024).

[372] Mike Wu and Noah Goodman. "Multimodal Generative Models for Scalable Weakly-Supervised Learning". In: *arXiv:1802.05335 [cs, stat]* (Nov. 2018). arXiv: 1802.05335. url: http://arxiv.org/abs/1802.05335 (visited on 03/16/2021).

[373] Zhengwang Wu, Gang Li, Li Wang, Feng Shi, Weili Lin, John H. Gilmore, and Dinggang Shen. "Registration-Free Infant Cortical Surface Parcellation Using Deep Convolutional Neural Networks". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger. Cham: Springer International Publishing, 2018, pp. 672–680. isbn: 978-3-030-00931-1. doi: 10.1007/978-3-030-00931-1_77.

[374] Zhengwang Wu, Fenqiang Zhao, Jing Xia, Li Wang, Weili Lin, John H. Gilmore, Gang Li, and Dinggang Shen. "Intrinsic Patch-Based Cortical Anatomical Parcellation Using Graph Convolutional Neural Network on Surface Manifold". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan. Cham: Springer International Publishing, 2019, pp. 492–500. isbn: 978-3-030-32248-9. doi: 10.1007/978-3-030-32248-9_55.

[375] Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. "Multi-VAE: Learning Disentangled View-common and View-peculiar Visual Representations for Multi-view Clustering". en. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9214–9223. isbn: 978-1-66542-812-5. doi: 10.1109/ICCV48922.2021.00910. url: https://ieeexplore.ieee.org/document/9710445/ (visited on 10/25/2023).

[376] Qingqing Yang, Xinxin Zhang, Yingchao Song, Feng Liu, Wen Qin, Chunshui Yu, and Meng Liang. "Stability test of canonical correlation analysis for studying brain-behavior relationships: The effects of subject-to-variable ratios and correlation strengths". en. In: *Human Brain Mapping* 42.8 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25373, pp. 2374–2392. issn: 1097-0193. doi: 10.1002/hbm.25373. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25373 (visited on 09/29/2023).

[377] Lakshmi N. Yatham. "Biomarkers for clinical use in psychiatry: where are we and will we ever get there?" en. In: *World Psychiatry* 22.2 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wps.21079, pp. 263–264. issn: 2051-5545. doi: 10.1002/wps.21079. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/wps.21079 (visited on 04/05/2024).

[378] Shelly Yin, Seok-Jun Hong, Adriana Di Martino, Michael P Milham, Bo-Yong Park, Oualid Benkarim, Richard A I Bethlehem, Boris C Bernhardt, and Casey Paquola. "Shared and distinct patterns of atypical cortical morphometry in children with autism and anxiety". In: *Cerebral Cortex* 32.20 (Oct. 2022), pp. 4565–4575. issn: 1047-3211. doi: 10.1093/cercor/bhab502. url: https://doi.org/10.1093/cercor/bhab502 (visited on 04/11/2024).

[379] Jinsung Yoon, Yao Zhang, James Jordon, and Mihaela van der Schaar. "VIME: Extending the Success of Self- and Semi-supervised Learning to Tabular Domain". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 11033–11043. url: https://proceedings.neurips.cc/paper/2020/hash/7d97667a3e056acab9aaf653807b4a03-Abstract.html (visited on 01/18/2024).

[380] Peter N. E. Young, Mar Estarellas, Emma Coomans, Meera Srikrishna, Helen Beaumont, Anne Maass, Ashwin V. Venkataraman, Rikki Lissaman, Daniel Jiménez, Matthew J. Betts, Eimear McGlinchey, David Berron, Antoinette O'Connor, Nick C. Fox, Joana B. Pereira, William Jagust, Stephen F. Carter, Ross W. Paterson, and Michael Schöll. "Imaging biomarkers in neurodegeneration: current and future practices". In: *Alzheimer's Research & Therapy* 12.1 (Apr. 2020), p. 49. issn: 1758-9193. doi: 10.1186/s13195-020-00612-7. url: https://doi.org/10.1186/s13195-020-00612-7 (visited on 04/05/2024).

[381] Wu Youyou, Yang Yang, and Brian Uzzi. "A discipline-wide investigation of the replicability of Psychology papers over the past two decades". In: *Proceedings of the National Academy of Sciences* 120.6 (Feb. 2023). Publisher: Proceedings of the National Academy of Sciences, e2208863120. doi: 10.1073/pnas.2208863120. url: https://www.pnas.org/doi/10.1073/pnas.2208863120 (visited on 03/12/2024).

[382] Mariam Zabihi, Dorothea L. Floris, Seyed Mostafa Kia, Thomas Wolfers, Julian Tillmann, Alberto Llera Arenas, Carolin Moessnang, Tobias Banaschewski, Rosemary Holt, Simon Baron-Cohen, Eva Loth, Tony Charman, Thomas Bourgeron, Declan Murphy, Christine Ecker, Jan K. Buitelaar, Christian F. Beckmann, and Andre Marquand.

"Fractionating autism based on neuroanatomical normative modeling". en. In: *Translational Psychiatry* 10.1 (Nov. 2020). Publisher: Nature Publishing Group, pp. 1–10. issn: 2158-3188. doi: 10.1038/s41398-020-01057-0. url: https://www.nature.com/articles/s41398-020-01057-0 (visited on 03/13/2024).

[383] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stephane Deny. "Barlow Twins: Self-Supervised Learning via Redundancy Reduction". en. In: *Proceedings of the 38th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, July 2021, pp. 12310–12320. url: https://proceedings.mlr.press/v139/zbontar21a.html (visited on 01/18/2024).

[384] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. "S4L: Self-Supervised Semi-Supervised Learning". en. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, Oct. 2019, pp. 1476–1485. isbn: 978-1-72814-803-8. doi: 10.1109/ICCV.2019.00156. url: https://ieeexplore.ieee.org/document/9010283/ (visited on 01/18/2024).

[385] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. "Understanding deep learning requires rethinking generalization". en. In: *5th International Conference on Learning Representations, {ICLR} 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. Nov. 2016. url: https://openreview.net/forum?id=Sy8gdB9xx (visited on 02/29/2024).

[386] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. "mixup: Beyond Empirical Risk Minimization". en. In: *6th International Conference on Learning Representations, {ICLR} 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. Feb. 2018. url: https://openreview.net/forum?id=r1Ddp1-Rb (visited on 01/26/2024).

[387] Richard Zhang, Phillip Isola, and Alexei A. Efros. "Colorful Image Colorization". en. In: *Computer Vision – ECCV 2016*. Ed. by Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 649–666. isbn: 978-3-319-46487-9. doi: 10.1007/978-3-319-46487-9_40.

[388] Xiaoyu Zhang, Jingqing Zhang, Kai Sun, Xian Yang, Chengliang Dai, and Yike Guo. "Integrated Multi-omics Analysis Using Variational Autoencoders: Application to Pancancer Classification". In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Nov. 2019, pp. 765–769. doi: 10.1109/BIBM47256.2019.8983228. url: https://ieeexplore.ieee.org/document/8983228 (visited on 03/26/2024).

[389] Yu Zhang, Peter Tiňo, Aleš Leonardis, and Ke Tang. "A Survey on Neural Network Interpretability". In: *IEEE Transactions on Emerging Topics in Computational Intelligence* 5.5 (Oct. 2021). arXiv:2012.14261 [cs], pp. 726–742. issn: 2471-285X. doi: 10.1109/TETCI.2021.3100641. url: http://arxiv.org/abs/2012.14261 (visited on 03/05/2024).

[390] Fenqiang Zhao, Zhengwang Wu, and Gang Li. "Deep learning in cortical surface-based neuroimage analysis: a systematic review". In: *Intelligent Medicine* 3.1 (Feb. 2023), pp. 46–58. issn: 2667-1026. doi: 10.1016/j.imed.2022.06.002. url: https://www.sciencedirect.com/science/article/pii/S2667102622000493 (visited on 02/21/2024).

[391] Fenqiang Zhao, Zhengwang Wu, Fan Wang, Weili Lin, Shunren Xia, Dinggang Shen, Li Wang, and Gang Li. "S3Reg: Superfast Spherical Surface Registration Based on Deep Learning". In: *IEEE Transactions on Medical Imaging* 40.8 (Aug. 2021). Conference Name: IEEE Transactions on Medical Imaging, pp. 1964–1976. issn: 1558-254X. doi: 10.1109/TMI.2021.3069645. url: https://ieeexplore.ieee.org/document/9389746 (visited on 04/03/2024).

[392] Fenqiang Zhao, Zhengwang Wu, Li Wang, Weili Lin, John H. Gilmore, Shunren Xia, Dinggang Shen, and Gang Li. "Spherical Deformable U-Net: Application to Cortical Surface Parcellation and Development Prediction". In: *IEEE Transactions on Medical Imaging* 40.4 (Apr. 2021). Conference Name: IEEE Transactions on Medical Imaging, pp. 1217–1228. issn: 1558-254X. doi: 10.1109/TMI.2021.3050072. url: https://ieeexplore.ieee.org/document/9316936 (visited on 01/18/2024).

[393] Fenqiang Zhao, Zhengwang Wu, Li Wang, Weili Lin, Shunren Xia, and Gang Li. "A Deep Network for Joint Registration and Parcellation of Cortical Surfaces". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Ed. by Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert. Cham: Springer International Publishing, 2021, pp. 171–181. isbn: 978-3-030-87202-1. doi: 10.1007/978-3-030-87202-1_17.

[394] Fenqiang Zhao, Shunren Xia, Zhengwang Wu, Dingna Duan, Li Wang, Weili Lin, John H. Gilmore, Dinggang Shen, and Gang Li. "Spherical U-Net on Cortical Surfaces: Methods and Applications". en. In: *Information Processing in Medical Imaging*. Ed. by Albert C. S. Chung, James C. Gee, Paul A. Yushkevich, and Siqi Bao. Cham: Springer International Publishing, 2019, pp. 855–866. isbn: 978-3-030-20351-1. doi: 10.1007/978-3-030-20351-1_67.

[395] Hui Zou and Trevor Hastie. "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67.2 (2005). Publisher: [Royal Statistical Society, Wiley], pp. 301–320. issn: 1369-7412. url: https://www.jstor.org/stable/3647580 (visited on 03/14/2024).

[396] Kaifeng Zou, Sylvain Faisan, Fabrice Heitz, and Sébastien Valette. "Joint Disentanglement of Labels and Their Features with VAE". In: *2022 IEEE International Conference on Image Processing (ICIP)*. ISSN: 2381-8549. Oct. 2022, pp. 1341–1345. doi: 10.1109/ICIP46576.2022.9898046. url: https://ieeexplore.ieee.org/abstract/document/9898046?casa_token=SPU9shFuy9QAAAAA:T66gUM9aMyJPcad6cvpRIHUeS4KwsVkxt5-nB4-4wwg6u7Zxpy3rtMzhAAOuIDuNcKpGq1jONzE (visited on 02/20/2024).