



Degree Project in Machine Learning

Second cycle, 30 credits

Learning deep representations of brain sMRI informed by latent psychopathology dimensions

PAULINE AMROUCHE

Learning deep representations of brain sMRI informed by latent psychopathology dimensions

PAULINE AMROUCHE

Degree Programme in Computer Science and Engineering

Date: March 19, 2026

Supervisors: Benoit Dufumier, Pawel Andrzej Herman

Examiner: Erik Fransén

School of Electrical Engineering and Computer Science

Host company: CEA

Host organization: NeuroSpin - UMR9027 Baobab

Swedish title: Psykiatriskt informerade MR-representationer med djupinlärning för klinisk prediktion

Abstract

Identifying neural correlates of psychiatric disorders is a fundamental step towards precision psychiatry – a paradigm aiming to discover biomarkers for early disease detection or personalised treatment selection to eventually improve patients’ quality of life. However, progress has been historically hindered by the classification of psychiatric disorders as discrete entities, suspected of not aligning with the disorders’ underlying biology. In response, modern frameworks have redefined psychiatric phenotypes along continuous latent psychopathology dimensions (LP-dimensions) aligned with biological and behavioural observations and therefore better suited for uncovering brain-behaviour correlations. Another limitation to this research is the small sample sizes of available clinical datasets. Deep representation learning offers a solution to this data scarcity by leveraging large-scale general datasets to learn transferable features to clinical cohorts. In this work, we investigate a method that combines LP-dimensions with transfer learning to generate clinically relevant representations of structural magnetic resonance imaging (sMRI) of the brain. We introduce *psy-Aware*, a contrastive learning model derived from *y-Aware* that incorporates LP-dimensions as weak supervision. We evaluate whether this approach improves upon standard contrastive learning using *SimCLR*. Our analyses focus on youth psychopathology, exploiting the large-scale, longitudinal ABCD study. This work makes three main contributions: (1) we derive a longitudinal three LP-dimensions model of youth psychopathology using exploratory factor analysis; (2) we benchmark two contrastive representation learning models on diagnosis prediction tasks; (3) we propose novel visualisations of the *y-Aware* loss and relate its behaviour with data distribution. We find that incorporating LP-dimensions yields comparable performances to standard contrastive learning approaches without significant improvement. We further identify the highly skewed distribution of LP-dimension as a likely limitation, suggesting that current psychiatric assessments may lack sufficient variance to fully support a dimensional approach in a broad population.

Keywords

Contrastive learning, Representation learning, Psychiatry, Magnetic Resonance Imaging, Medical imaging, Neuroscience

Sammanfattning

Att identifiera neurala korrelat för psykiatriska tillstånd är ett grundläggande steg mot precisionspsykiatri – ett paradigm som syftar till att upptäcka biomarkörer för tidig diagnostik eller personcentrerade behandlingsval, för att i slutändan förbättra patienters livskvalitet. Historiskt sett har framstegen dock hämmats av att psykiatriska diagnoser klassificerats som diskreta enheter, vilket misstänks sakna överensstämmelse med tillståndens underliggande biologi. Som ett svar på detta har moderna ramverk omdefinierat psykiatriska fenotyper som kontinuerliga dimensioner. Dessa är förankrade i biologiska och beteendemässiga observationer och är därmed bättre lämpade för att identifiera korrelationer mellan hjärna och beteende. En annan begränsning för denna forskning är den ringa storleken på tillgängliga kliniska dataset. Representationsinlärning erbjuder en lösning på denna databrist genom att utnyttja storskaliga, generella dataset för att lära sig överförbara särdrag för kliniska kohorter. I detta arbete undersöker vi en metod som kombinerar ett dimensionellt förhållningssätt till psykiatri med överföringsinlärning för att generera kliniskt relevanta representationer av strukturell magnetresonanstomografi av hjärnan. Vi introducerar *psy-Aware*, en kontrastiv inlärningsmodell som bygger på *y-Aware* och som integrerar psykopatologiska faktorer som svag övervakning under träning. Vi utvärderar huruvida detta tillvägagångssätt förbättrar standardmetoder för kontrastiv inlärning, såsom *SimCLR*. Analyserna fokuserar på psykopatologi hos barn och ungdomar och utnyttjar den storskaliga, longitudinella ABCD-studien. Arbetet ger tre huvudsakliga bidrag: (1) härledning av en longitudinell trefaktormodell för ungdomars psykopatologi med hjälp av explorativ faktoranalys; (2) jämförande utvärdering av två kontrastiva representationsinlärningsmodeller på diagnostiska prediktionsuppgifter; samt (3) introduktion av nya visualiseringar av *y-Aware*-förlustfunktionen och en analys av dess beteende i relation till datadistributionen. Våra resultat visar att integrering av psykopatologiska dimensioner ger jämförbar prestanda med standardiserad kontrastiv inlärning, utan signifikant förbättring. Vidare identifierar vi den starkt snedfördelade fördelningen av latent psykopatologiska faktorer som en sannolik begränsning, vilket tyder på att nuvarande psykiatriska bedömningsinstrument kan sakna tillräcklig varians för att fullt ut stödja ett dimensionellt angreppssätt i breda populationsstudier.

Nyckelord

Kontrastiv inlärning, Representationsinlärning, Psykiatri, Magnetisk resonanstomografi, Medicinsk bildvetenskap, Neurovetenskap

Acknowledgments

I would like to warmly thank my supervisor at Neurospin, Benoit Dufumier, for his thoughtful guidance and enlightening intuitions and insights on contrastive learning. I am thankful to Edouard Duchesnay, my second supervisor and head of GAIA lab, for his trust and for introducing me to the world of psychiatric neuroimaging.

I also want to thank Pawel Herman and Erik Fransén, my supervisor and examiner at KTH, for their time and valuable feedback on writing the thesis milestones.

Finally, I extend my thanks to my colleagues at the GAIA lab for making this thesis both a scientifically stimulating and enjoyable journey.

This work received funding via the research program in precision psychiatry (PEPR PROPSY, ANR-22-EXPR-0001), both funded by the France 2030 program and the French National Research Agency (ANR). This work was performed using HPC resources from GENCI-IDRIS (Grant 2024-AD011015990). The ABCD data used in this report were downloaded from the NIMH Data Archive under Data Use Certification #19095. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or ABCD consortium.

Saclay, France, March 2026

Pauline Amrouche

Contents

1	Introduction	1
1.1	Background	1
1.2	Problem statement	3
1.2.1	Research question	3
1.2.2	Objectives	3
1.3	Aim and scope	4
1.4	Thesis outline	6
2	Background	7
2.1	Dimensional psychiatric phenotyping	7
2.1.1	Dimensional paradigms	7
2.1.2	Assessment instruments in youth psychopathology	7
2.1.3	Factor analysis for youth psychopathology	8
2.1.3.1	Delineating latent dimensions of youth psychopathology	8
2.1.3.2	Exploratory factor analysis	9
2.1.3.3	Methodological considerations and limitations	10
2.2	Neuroimaging data in youth for the study of psychopathology	11
2.2.1	Cohorts and psychopathology measurement	11
2.2.2	From features to whole-brain analysis	12
2.2.3	From statistical towards deep learning correlation methods	12
2.2.4	Key findings	13
2.2.5	Limitations and gaps	13
2.2.6	Predicting dimensions of psychopathology from neuroimaging	13
2.3	Transfer learning and contrastive learning for neuroimaging	14
2.3.1	The curse of dimensionality in neuroimaging	14

2.3.2	Transfer learning to overcome the curse of dimensionality	14
2.3.3	Contrastive learning to learn transferable imaging representation	15
2.3.4	SimCLR: an unsupervised contrastive learning framework	16
2.3.5	y-Aware: a weakly-supervised contrastive learning framework	17
2.3.6	Data augmentation in 3D neuroimaging	17
3	Methods	19
3.1	Dimensions of youth psychopathology	19
3.1.1	Participants	19
3.1.2	Behavioural data	19
3.1.3	Exploratory factor analysis	20
3.1.4	Analysis of resulting dimensional scales	21
3.2	Data and preprocessing for machine learning	21
3.2.1	Datasets	21
3.2.1.1	ABCD Dataset	21
3.2.2	Clinical downstream datasets	22
3.2.3	MRI preprocessing	22
3.3	Machine Learning predictive modelling	23
3.3.1	Ridge regression baseline	23
3.3.2	Self-supervised contrastive learning	23
3.3.2.1	Architectures	23
3.3.2.2	Hyperparameters choice	24
3.4	Downstream tasks and evaluation	25
3.4.1	Downstream tasks	25
3.4.2	Statistical testing	26
3.5	y-Aware loss visualisations	26
3.6	Implementation	27
4	Results	29
4.1	CBCL-derived continuous scales of psychopathology	29
4.1.1	On baseline data, our analysis reproduces the results from Michelini et al.'s study	29
4.1.2	The three-factor structure is stable across time	31
4.2	Externalising and neurodevelopmental dimensions are better predicted from VBM than the internalising symptoms	32

4.3	Incorporating dimensions of psychopathology yields comparable performances to standard contrastive learning	33
4.3.1	Hyperparameters search	34
4.4	The highly skewed factor distribution may deteriorate training quality	35
5	Discussion and limitations	39
5.1	Factor analysis is restricted to young teenagers	40
5.2	Skewed metadata distribution limits the application of y -Aware	40
5.3	Low resolution of latent psychopathology dimensions in healthy subjects	40
6	Conclusions and Future work	43
6.1	Conclusions	43
6.2	Future works	44
6.3	Reflections	45
	References	47
A	Methods details	57
A.1	Clinical datasets demographics	57
A.2	Child Behavior Checklist (CBCL) items for exploratory factor analysis	57
A.3	Data augmentation parameters	59
B	Additional results	61

List of Figures

1.1	Proposed methodology	5
4.1	EFA loadings across time points	30
4.2	Linear prediction of latent psychopathology dimensions from ROI features	32
4.3	Downstream diagnosis classification results	33
4.4	UMAP projections of embedding space	36
4.5	Weights matrix from y-Aware loss	37
4.6	Parallel between distribution of neighbours in a batch and distribution of metadata	38
A.1	Data augmentations	60

List of Tables

3.1	ABCD demographics for EFA	20
3.2	Demographics of ABCD data	22
4.1	Stability of loadings across time points	31
4.2	Hyperparameters search for data augmentation	34
A.1	Demographics of clinical datasets	58
B.1	Loadings of the longitudinal 3-factor model	61

List of acronyms and abbreviations

ABCD	Adolescent Brain Cognitive Development Study
ASD	Autism Spectrum Disorder
BD	Bipolar Disorder
CBCL	Child Behavior Checklist
EFA	Exploratory Factor Analysis
LP- dimensions	Latent Psychopathology dimensions
MRI	Magnetic Resonance Imaging
ROI	Region of Interest
SCZ	Schizophrenia
sMRI	Structural Magnetic Resonance Imaging
VBM	Voxel-Based Morphometry

Chapter 1

Introduction

1.1 Background

In the early 20th century, neurology and psychiatry diverged, creating a “great divide” (Price et al., 2000) between the two disciplines based on a brain/mind dichotomy hypothesis. Neurological disorders were characterized by observable and measurable anatomical or functional brain abnormalities, while psychiatric disorders analyses relied on more subjective behavioural symptoms with no physical brain pathology (Kandel, 1999; Price et al., 2000). However, this hypothesis has been challenged by biological psychiatrists since its origins, and the rise of transdisciplinary research at the turn of the 21st century (Kandel, 1999) has fostered a growing body of work aimed at **uncovering the neural correlates of psychopathology**. This research seeks to deepen our understanding of the neurobiology underlying psychiatric disorders and provide clinicians with objective, measurable biomarkers of disorders. This would complement traditional behavioural assessments and enhance early diagnosis and personalised treatment, an approach coined as **precision psychiatry** (Bzdok & Meyer-Lindenberg, 2018; Fernandes et al., 2017; Segal et al., 2025). This effort towards precision psychiatry also involves the development of machine learning models able to extract the complex patterns of brain-behaviour correlations from data (Bzdok & Meyer-Lindenberg, 2018; Fernandes et al., 2017).

Recent advances in representation learning have provided new tools for identifying such correlations. Previous works demonstrated that **self-supervised contrastive methods** can generate semantically rich **representations of brain anatomy**, enabling predictions of age, biological sex, or neurodegenerative diagnoses (Dufumier et al., 2024; Kaczmarek et al., 2025).

Applying a similar approach to psychiatric phenotypes could yield clinically informative representations of brain anatomy, serving as a crucial step toward identifying neural correlates. Additionally, adolescence represents a critical developmental stage, as over 60% of mental health disorders emerge before the age of 25 (Solmi et al., 2022; Uhlhaas et al., 2023), making the study of psychopathology during this period particularly significant.

Previous studies have faced limitations in two main aspects: the precision and adequacy of psychiatric representations (Insel et al., 2010; Kotov et al., 2017; Segal et al., 2025; Tiego et al., 2023) and the sample size and resolution of neuroanatomical imaging data (Dhamala et al., 2023; Marek et al., 2022; Varoquaux, 2018).

A recent conceptual shift in psychiatry aimed to address the first limitation on the definition of the psychiatric phenotype. Traditionally, psychiatric disorders have been classified using categorical frameworks, such as the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), which define discrete diagnostic entities based on symptom clusters. While influential in both clinical and research settings, this approach has been increasingly criticized for its limited specificity, substantial symptom overlap across diagnoses, and high rates of comorbidity (Insel et al., 2010; Kotov et al., 2017; Segal et al., 2025; Tiego et al., 2023). In response, alternative frameworks have emerged that conceptualize psychopathology along continuous, hierarchical dimensions rather than rigid categories. This **dimensional approach of psychiatry** may align more closely with the observed symptoms and offers a promising path for investigating underlying neurobiological correlates (Tiego et al., 2023). Previous studies (Achenbach, 2001; Michelini et al., 2019) have employed **factor analytic methods** to identify **latent psychopathology dimensions (LP-dimensions)** in youth.

In parallel, different initiatives in the neuroimaging community led to the development of large cohorts combining behavioural and imaging data to overcome the second limitation. One notable example is the **Adolescent Brain Cognitive DevelopmentSM (ABCD) study** (“ABCD Study,” n.d.), which has compiled one of the largest youth brain dataset including over 10,000 participants scanned at multiple time points. By integrating neuroanatomical scans with mental health assessments, this dataset provides a rich resource for advancing research in youth psychopathology.

Finally **transfer learning** has enabled the use of such general large-scale datasets as ABCD to learn features of brain anatomy that can then be transferred for clinical predictive tasks in smaller cohorts with specific psychiatric phenotypes (Caruana, 1997; Dufumier et al., 2024). Additionally

while most prior studies have relied on feature-based representations of imaging data, using anatomical features such as cortical thickness or surface area averaged across regions of interest (ROIs) to produce compact representations. The scale of available datasets now enables the direct use of **voxel-resolution images** (approximately 1.5 mm³ isotropic), such as those obtained through voxel-based morphometry (VBM), an automatic estimation of grey matter tissue volumes from the raw MRI scan. Dufumier et al. (2021) introduced a model that generates such voxel-level representation. Building on the popular SimCLR framework (Chen et al., 2020), they defined the **y-Aware weakly supervised loss**, which incorporates auxiliary information about samples to guide the learning process. For instance, age has been used to refine representations of brain anatomy. This raises the question of whether clinical auxiliary data could further enhance these representations.

1.2 Problem statement

As a contribution to the larger effort towards precision psychiatry for youth, this master's thesis investigates a way to overcome the current limitations to uncovering neural correlates of psychiatric disorders. To do so, this work integrates three recent advances described in the previous paragraph in order to produce representations of neuroanatomy that correlate with psychiatric phenotype:

- Defining the psychiatric phenotype along latent psychopathology dimensions,
- The large-scale ABCD dataset of youth imaging and behavioural data,
- The y-Aware weakly supervised contrastive learning framework to incorporate auxiliary metadata in training.

1.2.1 Research question

The main question of this work is: What is the effect of incorporating latent psychopathology dimensions into contrastive learning with VBM imaging on the quality of youth brain anatomy representations for clinical applications?

1.2.2 Objectives

Figure 1.1 presents a visual summary of our proposed approach. Our research question integrates three key elements: defining psychiatric phenotypes along

LP-dimensions (Figure 1.1A), utilising VBM as the neuroimaging input, and employing weakly-supervised contrastive learning as the modelling approach. To systematically address this question, we structure our work into two sequential sub-objectives:

1. We build upon previous work by Michelini et al. (2019), who identified LP-dimensions in youth (Figure 1.1A) using a factor analysis method on the ABCD dataset at baseline (ages 9-10) (Figure 1.1B). we aim to assess the temporal stability of these factors during early adolescence using longitudinal ABCD data and derive longitudinal LP-dimensions.
2. As a sanity check, we evaluate the potential of using VBM imaging as an adequate preprocessing for finding correlations of neuroanatomy with LP-dimensions.
3. We introduce the psy-Aware loss incorporating LP-dimensions in a weakly-supervised contrastive learning loss (Figure 1.1C). We investigate our research question by comparing the proposed method with traditional fully-unsupervised contrastive learning. We evaluate all models on downstream psychiatry diagnosis tasks to assess their potential for clinical analyses (Figure 1.1D). To deepen our analysis, we also study how the integration of multiple LP-dimensions influences contrastive learning model training.

1.3 Aim and scope

This work aims to contribute to the broader international collaboration for precision psychiatry (Fernandes et al., 2017) by investigating a promising representation learning approach to learn neuroanatomy representations that enable clinical predictions.

In particular in the search for neural correlates of psychiatric disorder, the scope of this thesis is limited to neuroanatomy as observed on a T1-weighted MRI scan. Therefore, we exclude from the analysis other modalities such as non-MRI signals, functional MRI (fMRI), diffusion weighted imaging (DWI) or other neuroanatomy contrasts like T2-weighted MRI or Fluid-attenuated inversion recovery (FLAIR). More restrictively because of time constraints we limit this analysis to imaging preprocessed using VBM. We focus on youth psychopathology with the use of the ABCD dataset (aged 9 to 15 years old) for pretraining. For the downstream tasks, we do not consider longitudinal tasks in this work.

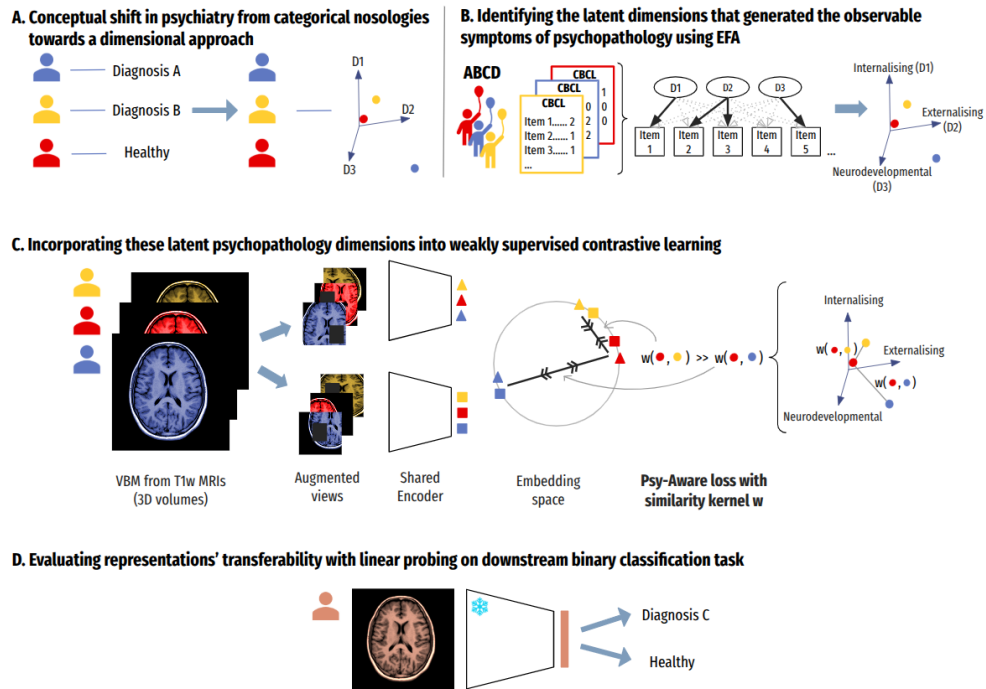


Figure 1.1: **Proposed methodology** — **A.** We adopted a dimensional approach of psychiatric phenotype defined from continuous latent psychopathology dimensions (LP-dimensions) (e.g. D1, D2, D3). **B.** We used exploratory factor analysis (EFA) on the Child Behaviour Checklist (CBCL) data to identify a 3 LP-dimensions model in the ABCD dataset. **C.** We then proposed the psy-Aware loss, incorporating the LP-dimensions as auxiliary variables in weakly supervised contrastive learning. The loss pulls closer the embeddings of images from participants with similar psychopathology scores. **D.** We evaluated the transferability of obtained representations to three diagnosis classification tasks: schizophrenia, bipolar disorder and autism spectrum disorder.

1.4 Thesis outline

The following chapters address our research question of whether integrating continuous latent psychopathology dimensions into weakly supervised contrastive learning can enhance representations of youth brain anatomy.

Chapter 2 reviews the relevant literature. Section 2.1 outlines the recent evolution of psychiatric frameworks and the shift towards dimensional approaches, and current assessment tools in youth, including prior work on delineating LP-dimensions using exploratory factor analysis. Section 2.2 reviews methodological approaches and key findings from studies investigating neural correlates of youth psychopathology using structural neuroimaging. Section 2.3 introduces deep learning concepts and challenges specific to neuroimaging, in particular it presents the SimCLR and y -Aware contrastive learning frameworks employed in this thesis.

Chapter 3 describes the study design and analyses, including data sources and participant demographics, preprocessing of behavioural and imaging data, derivation of longitudinal LP-dimensions, deep learning architectures, and downstream evaluation procedures.

Chapter 4 presents the empirical findings. It first reports the CBCL-derived longitudinal three LP-dimensions model and its temporal stability, then presents the sanity check assessing associations between VBM features and LP-dimensions, and finally evaluates the performance of the psy -Aware framework, including visual analyses of the learnt embedding space.

Chapter 5 interprets these results in relation to the research question, discussing the impact of skewed metadata distributions, limited resolution of psychopathology measures, and domain gaps on the effectiveness and transferability of y -Aware representations.

Finally, Chapter 6 summarises the main contributions, outlines study limitations, and proposes directions for future research.

Chapter 2

Background

2.1 Dimensional psychiatric phenotyping

2.1.1 Dimensional paradigms

As presented in the introduction, pluridisciplinary research in the 21st century led to the development of dimensional frameworks of psychiatry to answer the limitations of categorical ontologies (Insel et al., 2010; Kandel, 1999; Kotov et al., 2017). The National Institute of Mental Health introduced in 2010 the Research Domain Criteria (RDoC) initiative (Insel et al., 2010), a theory-driven, construct-based approach that organises psychopathology across multiple domains of functioning. A parallel initiative, the Hierarchical Taxonomy of Psychopathology (HiTOP) (Kotov et al., 2017) adopts a data-driven, bottom-up perspective, inferring latent dimensions through statistical analysis of symptom patterns in large populations (Tiego et al., 2023). Dimensional models better capture clinical heterogeneity and provide a refined framework for studying neurobiological correlates of psychopathology (Segal et al., 2025).

2.1.2 Assessment instruments in youth psychopathology

The first versions of the DSM provided limited categorical descriptions of youth psychopathology, leading to the early development of dimensional assessment tools and their current use in common clinical practice (Achenbach, 2020). Among these, the Achenbach System of Empirically Based Assessment (ASEBA) stands as one of the most extensively adopted systems, originally

conceptualised in the 1970s with its first version published in the 1980s and later refined in subsequent updates. The current ASEBA system for children aged 6-18 includes three primary instruments: the Child Behavior Checklist (CBCL), Youth Self-Report (YSR), and Teacher Report Form (TRF), each targeting different informants. Alternative instruments include the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) and its computerised adaptation GOASSESS (Calkins et al., 2015). Because of the large pre-existing literature using the CBCL as a measure of psychopathology, we chose this instrument in our study.

The CBCL is a caregiver-report questionnaire designed for youth aged 6–18 years. The version used in this study, revised in 2001 (Achenbach, 2001), contains 118 items rated on a 3-point Likert scale (0=not true, 1=somewhat or sometimes true, 2=very true or often true). These items encompass a broad range of behavioural, emotional, and social functioning domains. The items are aggregated into eight empirically derived symptom dimensions: anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviour, and aggressive behaviour. These scales are further grouped into two broadband dimensions: internalising problems (inward-directed difficulties like anxiety or withdrawal) and externalising problems (outward-directed behaviours like aggression, rule-breaking, and hyperactivity). Raw CBCL scores are typically converted to age- and gender-normed T-scores based on United States population norms.

2.1.3 Factor analysis for youth psychopathology

2.1.3.1 Delineating latent dimensions of youth psychopathology

Factor analysis is a statistical method that tries to identify latent generating factors from a set of observable items. In the case of psychopathology, it relies on the assumption that the observed variability in individual youth symptoms (here the CBCL items) can be explained by a set of common higher-order factors. When one generating factor is considered, the resulting scale is termed "p-factor" and represents a general burden of psychopathology shared across pathology. For harmonisation across this manuscript, we call these factors latent psychopathology dimensions (LP-dimensions).

The CBCL includes empirically derived symptom dimensions, originally constructed by Achenbach and Rescorla (Achenbach, 2001) using factor analyses across large normative samples (e.g., CBCL: $n = 4,994$; YSR: $n = 2,581$; TRF: $n = 4,437$). These analyses yielded the eight-factors model

of child psychopathology described previously. However, this seminal work is lacking some methodological details to be fully reproducible. Although factor analytic methods (exploratory and confirmatory factor analysis on tetrachoric matrices) were reportedly used, the choice of the number of factors and the model integration across questionnaires are not fully documented.

Michelini et al. (2019) derived a hierarchical structure from the CBCL using EFA in the ABCD study, identifying five scales with one to five latent factors and a hierarchical structure linking these different scales. They tried to derive structures with a higher number of factors but found them to be little interpretable with some factors having less than three items loading on them. In particular, they did not identify a similar structure to Achenbach's eight symptoms scale. The following subsections describe the EFA method and limitations in psychopathology.

2.1.3.2 Exploratory factor analysis

As exploratory factor analysis (EFA) is not the core of this thesis, a brief intuitive overview of the method is given. This subsection is highly inspired by chapter 16 of Cosma Shalizi's very didactical textbook (Shalizi, 2025) on data analysis. We refer the curious reader to this reference for a more detailed introduction to the subject.

Given a set $X \in \mathbb{R}^{n \times p}$ of the answers from n participants to p CBCL items, EFA tries to find a set of q latent factors explaining these observations. Each factor is described by a p -dimensional vector representing its loadings (or weights) on the individual CBCL items. The stacked vectors form the matrix $w \in \mathbb{R}^{q \times p}$. We note $F \in \mathbb{R}^{n \times q}$ the projection of the observations for each participant onto each factor. EFA thus approximates:

$$X \approx Fw$$

X is the observed data, F is the matrix of (latent, unobserved) factor scores and w the matrix of factor loadings. The goal is to define w to be able to compute F , the scores of each participant on the q factors of psychopathology. The core aim of EFA is to preserve **correlations** or covariance in the data. As opposed to principal components analysis which preserves variance. Mathematically, in its simplest version (with additional hypotheses) EFA derives w such that the empirical correlation matrix v of X is approximated by

$$v \approx w^T w$$

The empirical correlation matrix can be derived from Pearson's correlation or polychoric correlation. Polychoric correlation is designed for ordinal data and considers two sets of answers to be sampled from thresholded latent bivariate normal distributions. Polychoric correlation coefficient is the maximum likelihood estimate of the correlation between those two latent distributions (Olsson, 1979).

2.1.3.3 Methodological considerations and limitations

Standard EFA procedures based on the Pearson's correlation matrix of data and maximum likelihood estimate make the assumption of normally distributed observations. However, the CBCL responses are ordinal (discrete and ordered) and highly skewed towards 0 in the general population as most children have few symptoms of psychopathology. Therefore for ordinal data, standard EFA procedures are inadequate. Current best practices recommend using polychoric correlation matrices and alternative extraction methods such as minimum residuals (MINRES) (Goretzko et al., 2021; Lorenzo-Seva & Ferrando, 2021). However, polychoric correlations assume that ordinal responses arise from thresholded underlying bivariate normal distributions, but this assumption may be violated in practice, potentially introducing even greater bias (Grønneberg & Foldnes, 2024). Moreover because polychoric correlation is an estimate of the actual correlation between the underlying distributions, it is more vulnerable to measurement noise (up to 50%) than Pearson correlation (Lorenzo-Seva & Ferrando, 2021) and the resulting correlation matrix may not be definite positive. This situation leads to unstable results of EFA and can be mitigated by removing items with no or low variability and merging items that are highly correlated (Lorenzo-Seva & Ferrando, 2021). This helps bringing the correlation matrix closer to full rank.

Despite these limitations, polychoric-based EFA remains the standard approach in the literature (Hoffmann et al., 2022; Michelini et al., 2019; Moore et al., 2020). Other approaches such as bifactor models have also been employed to isolate general versus domain-specific factors (Hoffmann et al., 2022; Moore et al., 2020). The choice of the modelling approach may not be so crucial though, as Achenbach et al. (2025) recently demonstrated high convergence among different methods of computing the p-factor in the one-factor case.

2.2 Neuroimaging data in youth for the study of psychopathology

The development and advances in MRI technology since the 1970s and the availability of clinical neuroimaging cohorts have enabled the systematic investigation of brain structural and functional correlates of psychiatric disorders (Segal et al., 2025). Initial studies primarily employed case-control designs to identify neuroanatomical differences associated with individual disorders such as schizophrenia (Ortiz-Gil et al., 2011), bipolar disorder (Delvecchio et al., 2012), and depression (Costafreda et al., 2009; Delvecchio et al., 2012).

Recent shifts in psychiatry paradigms advocate for transdiagnostic and dimensional frameworks (see section 2.1) which better capture the complexity and comorbidity of psychiatric symptoms (Vanes & Dolan, 2021). A notable meta-analysis by Goodkind et al. demonstrated convergent grey matter reductions across six major psychiatric diagnoses, further motivating a transdiagnostic focus (Goodkind et al., 2015). Vanes and Dolan (2021) reviewed studies on neural markers of an unidimensional factor of psychopathology (termed p-factor). This section provides an updated overview of recent efforts to identify neural correlates of youth psychopathology.

2.2.1 Cohorts and psychopathology measurement

Three large-scale cohorts dominate the literature in youth psychopathology. The ABCD study follows about 12,000 American children from age 9-10 at baseline until age 14-15 in release 5 used in this thesis. Other prominent cohorts include Generation R (GenR) a Dutch longitudinal cohort from birth to adolescence, and the Philadelphia Neurodevelopmental Cohort (PNC) a cross-sectional sample of about 9,000 American children. Recently the Reproducible Brain Charts (RBC) initiative combined part of the PNC and 4 other developmental datasets into a harmonised, fully open and reproducible data resource to study brain imaging and mental health with more than 6,000 participants (Shafiei et al., 2025).

Psychopathology is typically assessed using the CBCL in ABCD and GenR, and the GOASSESS instrument in the PNC. Dimensional structures are generally derived from factor analysis or bifactor models. Several studies converge on a three dimensions structure encompassing internalising symptoms, externalising symptoms, and either an emotion dysregulation

(Blok et al., 2022), neurodevelopmental (Royer et al., 2024), or thought disorder component (Mewton et al., 2022). The Reproducible Brain Charts initiative developed a harmonised dimensional framework to incorporate data extracted from the CBCL and GOASSESS instruments into an unified bifactor model of psychopathology (Hoffmann et al., 2023).

2.2.2 From features to whole-brain analysis

With generally small datasets and high-dimensional inputs, brain imaging analysis is prone to the curse of dimensionality. To reduce the number of input features, many studies applied conventional statistical and machine learning techniques to region-based features extracted via brain parcellations, such as the Desikan-Killiany atlas (Blok et al., 2022; Mewton et al., 2022), or the Schaefer atlas (Parkes et al., 2021). Dimensionality reduction methods like principal component analysis (PCA) were also used to mitigate the high-dimensional nature of imaging data (Royer et al., 2024). Although feature extraction enhances interpretability, decreases noise and reduces computational burden, it may destroy valuable signals from the input (Dhamala et al., 2023). More recent work employed voxel-level analyses to retain spatial resolution in neuroanatomy input for correlation with autism spectrum traits (Aglinskas et al., 2022) or schizophrenia-related dimensions (Kirschner et al., 2020). Kaczkurkin et al. (2019) analysed volumetric T1-weighted MRI data and cortical thickness in a transdiagnostic framework using the Philadelphia Neurodevelopmental Cohort. While this review focuses on structural imaging, it is worth noting that functional MRI has shown stronger performance in predicting behavioural outcomes, likely due to its ability to capture dynamic neural processes (Dhamala et al., 2023; Ooi et al., 2022).

2.2.3 From statistical towards deep learning correlation methods

Most reviewed studies used multivariate analysis such as Partial Least Squares (PLS) (Kirschner et al., 2020; Royer et al., 2024) or Sparse Canonical Correlation Analysis (SCCA) (Xu et al., 2025) on anatomical features. Instead of using raw features, Parkes et al. (2021) applied non-linear normative modelling to yield deviation from normative features. Using those deviations instead of raw values of cortical grey matter volumes led to better prediction of transdiagnostic symptoms with a ridge regressor. However, the neural signature of disorders may be embedded in complex non-linear patterns, that

these methods cannot model (Abrol et al., 2020). Deep learning has shown promise in the field. For example, Aglinskias et al. (2022) applied contrastive variational autoencoder for autism-specific representation learning. But despite their promise, the predictive power of the machine learning models remains modest. In the ABCD cohort, Hill et al. found that the inclusion of imaging features did not improve prediction of future risk status beyond symptom data, highlighting current limitations in imaging-derived biomarkers (Hill et al., 2025).

2.2.4 Key findings

Many studies report transdiagnostic reductions in overall grey matter volume (Goodkind et al., 2015; Mewton et al., 2022; Parkes et al., 2021; Romer et al., 2023; Xu et al., 2025) and reductions in brain surface area (Mewton et al., 2022; Romer et al., 2023; Xu et al., 2025). Conversely, associations with cortical thickness are inconsistent as Romer et al. (2023) and Mewton et al. (2022) found none while Royer et al. (2024) reported reduced thickness linked to internalising symptoms.

2.2.5 Limitations and gaps

Findings obtained with brain imaging often lack replicability, and predictive accuracy remains low (Woo et al., 2017). This may reflect overfitting or the inadequacy of current models to capture the high heterogeneity intrinsic to psychiatric phenotypes. Traditional case-control designs risk oversimplification by assuming a single brain-behaviour mapping, potentially obscuring subgroup-specific patterns (Zhao et al., 2025).

2.2.6 Predicting dimensions of psychopathology from neuroimaging

Some previous works have tried to predict dimensions of psychopathology from MRI-derived neuroimaging data. For instance Ooi et al. (2022) have identified significantly better predictive power from functional data than from anatomical features to predict the a CBCL-derived measure of psychopathology in the ABCD dataset with kernel ridge regression. They extracted cortical volume, cortical thickness and cortical area features from the Schaefer 400 atlas (Schaefer et al., 2018), resulting in 1200 features. They find a correlation of 0.06 between the predicted and true values for the measure of

psychopathology using functional connectivity, and of 0.02 using anatomical features.

2.3 Transfer learning and contrastive learning for neuroimaging

2.3.1 The curse of dimensionality in neuroimaging

MRI data is typically represented as 3D volumes with around 10^6 voxels at 1.5 mm^3 resolution. In contrast, even the largest neuroimaging cohorts contain only up to 100,000 subjects (“UK Biobank,” 2025). For reference ImageNet in naturalistic images contains more than 14 million images (Russakovsky et al., 2015). This disproportion between data dimensionality and sample size makes deep neural networks in medical imaging highly prone to overfitting. Data reduction strategies, such as downsampling or parcellation into predefined ROIs, can alleviate this issue, but they also risk discarding potentially informative anatomical detail (Dhamala et al., 2023). Conversely, higher-resolution data preserve fine-grained information but require models that are sufficiently expressive to capture variability, while still employing strong regularization to avoid overfitting.

The use of machine learning (ML) and deep learning (DL) in neuroimaging requires careful consideration of sample size. Due to the high dimensionality of imaging data and the relatively small effect sizes in psychiatric populations, there is a risk for underpowered studies, inflated performance estimates and poor generalization (Dhamala et al., 2023; Marek et al., 2022; Varoquaux, 2018). Transfer learning offers a promising solution: pretraining on large-scale population datasets (e.g., UK Biobank, ABCD) and fine-tuning on smaller clinical cohorts has been shown to improve robustness and generalizability. For example, Dufumier et al. (2024) and Chopra et al. (2024) demonstrated that pretrained models outperform those trained directly on small clinical samples.

2.3.2 Transfer learning to overcome the curse of dimensionality

The general idea of transfer learning is to use a model already trained on one task for another task, under the hypothesis that it learned in the first task transferable features that will be useful for the second task (Caruana, 1997).

A common way to implement transfer learning is to use an encoder-projection head architecture. The encoder part – also called a backbone – maps the input into an embedding space learning a relevant representation for the task, this is the transferable component. The projection head then comes after this representation to solve the task, hence it is task-specific and needs to be trained for each task. The first task used to train the encoder is called the pretraining task while the subsequent task is called the downstream task. After transfer, the encoder can be further trained to adapt to the new task, this is called finetuning. When the encoder is transferred without further training and only a linear head is trained, this is called linear evaluation or linear probing.

In neuroimaging, transfer learning has notably been used (Dufumier et al., 2024) to pretrain a model on large datasets, more than a thousand subjects, of brain scans from the general population before transferring them to small clinical datasets, typically a few hundred subjects with specific diagnoses.

The value of transfer learning lies in the embedding space – the learnt representation of the input. The representation is considered useful if it aligns with the objectives of the downstream tasks, i.e. if the tasks become easier to solve using this representation. One way of assessing representation quality is linear probing. If the downstream tasks can be solved with a simple linear model then it means that the embedding space is organised in a relevant way for this task. For example, if brain scans are embedded in a 2D space such that scans of subjects with bipolar disorder have a negative x-coordinate and those from healthy subjects have a positive x-coordinate, then a very simple linear classifier (boundary $x = 0$) would solve the task of identifying subjects with bipolar disorder. However this same representation may be poorly suited for predicting age. Therefore the evaluation of representation quality is task-dependent.

2.3.3 Contrastive learning to learn transferable imaging representation

For classification downstream tasks, it is desirable for the embedding space to be organised according to semantic content. For example, in natural images, pictures of dogs clustered together, pictures of cats clustered near dogs but further from pictures of pastries. Contrastive learning is built on this intuition: similar inputs should be mapped close together, while dissimilar ones should be pushed apart.

Formally, contrastive learning is based on a contrastive loss between training examples. The loss penalizes similar examples being far apart and

dissimilar examples being close. Historically, Noise-Contrastive Estimation (NCE) (Gutmann & Hyvärinen, 2010) was introduced as a statistical method for density estimation by contrasting data against noise. This idea was later refined into the InfoNCE loss (Oord et al., 2019), which has become a reference in modern self-supervised representation learning. In 2020, Chen et al. introduced SimCLR (Chen et al., 2020), a simple self-supervised contrastive learning framework that achieved state-of-the-art performance on ImageNet and triggered a surge of work in self-supervised learning. Of these works, the y-Aware approach (Dufumier et al., 2021), leverages weakly-supervised learning to incorporate auxiliary labels into the embedding.

2.3.4 SimCLR: an unsupervised contrastive learning framework

SimCLR (Chen et al., 2020) relies on the InfoNCE loss (Oord et al., 2019) between positive (similar) and negative (dissimilar) pairs. Within each batch, two augmented views of the same image form a positive pair, while all other views serve as negatives. Data augmentations include rotations, cropping, noise injection, and color distortions. As this framework does not require labels, SimCLR is called a self-supervised technique. The loss function, named normalised temperature-scaled cross entropy loss (NT-Xent), derived from InfoNCE is defined as:

$$\mathcal{L}_{\text{NT-Xent}} = -\frac{1}{2N} \sum_{i=1}^{2N} \log \frac{\exp(\text{sim}(z_i, z_{i+N})/\tau)}{\sum_{\substack{k=1 \\ k \neq i}}^{2N} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (2.1)$$

where z_i and z_{i+N} are the embeddings from two views of the same image, z_k are the augmented views from all the other images in the batch, N is the number of samples in the batch, $\text{sim}(\cdot, \cdot)$ denotes cosine similarity, and τ is a temperature parameter controlling distribution sharpness. Intuitively, the numerator increases the similarity between positive pairs, while the denominator contrasts them against all other samples in the batch, treated as negatives. Minimizing this loss organises the embedding space such that semantically similar instances are clustered together and dissimilar ones are separated.

The SimCLR architecture consists of three main components: (i) a base encoder (e.g., ResNet) that extracts representations, (ii) a projection head (a small MLP) that maps these representations into a latent space where contrastive loss is applied, and (iii) a contrastive objective that enforces

similarity between positive pairs and dissimilarity between negative pairs. This separation of encoder and projection head has been shown to improve representation quality.

2.3.5 y-Aware: a weakly-supervised contrastive learning framework

In SimCLR, all other images in the batch are treated as negatives, although not all are equally dissimilar (e.g., cats and pastries are more dissimilar than cats and dogs). In a setting where the dissimilarity is continuous, it can be incorporated into training through the y-Aware loss (Dufumier et al., 2021). For example, neuroanatomy evolves with age and therefore in contrastive learning it makes sense to consider scans of two young people as more similar than scans from two persons with a larger age gap. In the original y-Aware paper, the authors incorporated age as an auxiliary variable in the loss and showed increased performances on several clinical tasks in psychiatry (Dufumier et al., 2021). Formally, y-Aware introduces a weighted contrastive loss:

$$\mathcal{L}_{\text{y-Aware}} = -\frac{1}{2N} \sum_{i=1}^{2N} \sum_{\substack{j=1 \\ j \neq i}}^{2N} \frac{w_{\Sigma}(y_i, y_j)}{\sum_{\substack{k=1 \\ k \neq i}}^{2N} w_{\Sigma}(y_i, y_k)} \log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{\substack{k=1 \\ k \neq i}}^{2N} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (2.2)$$

where z_i and z_{i+N} are the embeddings from two views of the same image i in the batch, y_i is the auxiliary label for image i , N is the number of samples in the batch, $\text{sim}(\cdot, \cdot)$ denotes cosine similarity, τ is a temperature parameter and w_{σ} is a gaussian kernel of bandwidth Σ . We recognize the NT-Xent loss in the second term of the product which is modulated by the first ratio term quantifying how similar the auxiliary labels are (the greater the ratio, the more these views will be considered as a soft positive pair). We name this ratio $weight(i, k) = \frac{w_{\Sigma}(y_i, y_j)}{\sum_{\substack{k=1 \\ k \neq i}}^{2N} w_{\Sigma}(y_i, y_k)}$ and refer to the matrix of pairwise weights between samples in a batch as the batch's weights matrix.

2.3.6 Data augmentation in 3D neuroimaging

Data augmentation is a key element of contrastive learning to create multiple informative views of each sample. In natural image datasets, this is achieved through strong transformations such as color jittering, blurring, rotations, or

flips, as in the seminal SimCLR framework (Chen et al., 2020). Translating these strategies to 3D MRI is non-trivial: what does it mean to rotate a brain volume, and how much can one alter a scan before destroying semantic information? Such design choices implicitly embed hypotheses about which features are relevant for the downstream task. For instance, Bashyam et al. (2020) used 2D CNNs with slice-based augmentations, while other work has emphasized 3D volume-preserving transformations. Dufumier et al. (2021) obtained good performances using only a random cutout of 1/16 of the image’s volume.

Recent approaches have also explored generative models for augmentation. For example, SynthSeg (Billot et al., 2023) synthesizes realistic segmentations from probabilistic priors. While promising, such methods are computationally expensive and not adopted here due to time constraints. Instead, we rely on non-generative 3D transformations (cropping, cut out and noise) implemented in dedicated libraries such as TorchIO (Pérez-García et al., 2021) and NIDL (“Nidl,” n.d.).

Chapter 3

Methods

3.1 Dimensions of youth psychopathology

3.1.1 Participants

We used data from the “ABCD Study” (n.d.), release 5, which includes longitudinal multimodal imaging and behavioural assessments of more than 10,000 American children collected across five time points: at baseline (ages 9–10) and then at four subsequent annual follow-ups (+1y, +2y, +3y and +4y). We did not use the fifth time point (+4y) for factor analysis as data was incomplete in release 5.

We selected participants with complete CBCL data at all time points resulting in n=6,290 unique participants. We first focused on reproducing the work presented in Michelini et al. (2019) based on the CBCL forms collected at baseline in the study. Then we extended this work by leveraging the subsequent data releases. The demographic details of participants available for each time point is presented in Table 3.1. Finally, we derived a longitudinal dimensional scale by leveraging the whole dataset. We randomly selected one time point among all available for each participant resulting in a sample incorporating data from all time points with youth from age 9 to 14, see last row of Table 3.1.

3.1.2 Behavioural data

In the ABCD dataset, psychopathology was assessed using the CBCL questionnaire for ages 6-18 (ASEBA system (Achenbach, 2001)), comprising 116 items rated on a 0-1-2 scale by the parents. To ensure that the items

	Age mean (std)	Biological sex	Number of subjects kept for longitudinal
Baseline	9.90 (0.62)	53%M, 47%F, 0.03%O	1553
+1y	10.91 (0.63)	53%M, 47%F, 0.03%O	1579
+2y	11.99 (0.65)	53%M, 47%F, 0.03%O	1509
+3y	12.90 (0.64)	53%M, 47%F, 0.03%O	1649
longitudinal	11.45 (1.30)	53%M, 47%F, 0.03%O	6290

Table 3.1: **ABCD demographics for EFA** — Age and sex demographics, $n=6,290$ for each time point (M:male, F:female, O:others)

correlation matrix is definite positive and to enhance factor stability (see section 2.1.3.3), Michelini et al. excluded items with no or low endorsement (less than 0.5% of the sample answering 1 or 2). However in line with Moore et al. (2020) we chose a slightly more exclusive threshold of 1% ($n = 19$ items excluded). We combined together items with high polychoric correlation ($> 75\%$) following similar procedures from Michelini et al. (Michelini et al., 2019) and Moore et al. (Moore et al., 2020). This yielded a total of 86 items including 10 composites. Full details are provided in Appendix A.2.

3.1.3 Exploratory factor analysis

We followed Michelini et al. (2019)'s methodology by using exploratory factor analysis to delineate higher-order factors of youth psychopathology from the CBCL items. To align with factor analysis literature, we use the denomination factors of psychopathology in this analysis. In the subsequent analyses, the term latent psychopathology dimensions (LP-dimensions) is used.

EFA computes loadings of the different CBCL items on a predetermined number of factors. The higher the loading, the more a particular item is representative of a factor. To enhance interpretability of the factor structure, we follow Michelini et al. (2019)'s criterion and we set a threshold on the loadings: if an item had a loading of more than 0.35 on a primary factor and differed by at least 0.10 from the second-highest loading then we retained this item's primary loading. This produced a loading matrix mapping observed items to latent factors where each item loads on at most one factor. Following best practices (Goretzko et al., 2021; Lorenzo-Seva & Ferrando, 2021), see 2.1.3.3, we applied factor analysis on the polychoric correlation matrix of the CBCL data, and EFA was conducted using minimum residual (MINRES)

extraction and geomin rotation. In this regard, we move slightly away from Michelini et al.'s methodology who used principal component analysis to extract factors.

3.1.4 Analysis of resulting dimensional scales

We compared the factors obtained on baseline data with the ones reported by Michelini et al. (2019) on the same dataset. We performed qualitative validation by visualising the loadings of the different scales. For quantitative validation we computed the correlation between the different scales as the mean absolute Pearson correlation coefficient between the individual unthresholded loadings of each factor across scales.

We found the three-factors scale to offer a good compromise between phenotyping precision and model complexity. Therefore we focused on this level of granularity for our subsequent analyses. We applied the qualitative and quantitative validations to assess the stability of loadings across all time points. Finally we used the longitudinal three-factors scale for the machine learning predictions.

3.2 Data and preprocessing for machine learning

3.2.1 Datasets

3.2.1.1 ABCD Dataset

We used T1-weighted (T1w) MRI scans for imaging and answers to the Child Behaviour Checklist (CBCL) for behavioural measures. We excluded participants with missing data in the CBCL or with no T1w MRI scan available resulting in a sample of 11,460 unique participants and $n=21,216$ scans across all five time points. For training the machine learning models we divided the full sample into three splits for training, validation and testing respectively. We split the sample based on the family identifier of participants, therefore ensuring that all scans from siblings or from the same participant at different time points are grouped within the same split. This is to avoid data leakage when training the model on a dataset that is too similar to the evaluation or testing datasets. Table 3.2 presents the demographics of each split.

Split	N scans	N participants	Age mean (std)	Biological Sex
Train	16,958	9,156	11.19 (1.60)	53%M, 47%F, 0.02%O
Validation	2,110	1,148	11.20 (1.59)	51%M, 49%F, 0.1%O
Test	2,148	1,156	11.17 (1.60)	53%M, 47%F, 0%O

Table 3.2: **Demographics of ABCD data** M:male, F:female, O:others

3.2.2 Clinical downstream datasets

For the clinical downstream tasks, we used the same datasets as in Dufumier et al.’s article (Dufumier et al., 2024) to evaluate our models on three tasks: classification of participants with bipolar disorder, schizophrenia and on the autism spectrum. Those datasets aggregate participants from the SCHIZCONNECT-VIP (“SchizConnect,” n.d.), CNP (Gorgolewski et al., 2017), PRAGUE, BSNIP (Tamminga, 2014), CANDI (“CANDI,” n.d.), BIOBD (Sarrazin et al., 2018), ABIDE1 (“ABIDE,” n.d.) and ABIDE2 (“ABIDE,” n.d.) studies. Table A.1 includes demographics of all datasets.

3.2.3 MRI preprocessing

T1-weighted MRI scans were preprocessed previously by the lab using CAT12 (Gaser et al., 2024) from the SPM toolbox. We used two different preprocessing of volumic grey matter, one at the region level and one at the voxel level yielding two granularities in imaging precision:

- **ROI features:** Regional grey matter volumetric measures were extracted after non-linear normalization to a standard template using the Neuromorphometrics atlas composed of 138 cortical and subcortical regions of interest.
- **VBM:** Structural scans were bias-field corrected, denoised, and segmented into grey matter, white matter, and cerebrospinal fluid compartments. Grey matter segments were then non-linearly warped onto the MNI template using the DARTEL algorithm (Ashburner, 2007) and modulated by the deformation’s Jacobian determinant to preserve local volumetric information. Images were then downsampled to an isotropic 1.5mm^3 spatial resolution. Total intracranial volume, estimated via CAT12, was used to normalise voxel intensities to account for individual differences in head size. This resulted in $113 \times 137 \times 113$ images for the ABCD dataset and in $121 \times 145 \times 121$ images for the

downstream datasets. The inconsistency is due to a different version of CAT12 being used for preprocessing. However as the images are then normalised to a 128x128x128 format we expect it to not affect the deep learning pipeline.

3.3 Machine Learning predictive modelling

3.3.1 Ridge regression baseline

As a linear baseline, Ridge regression (Tikhonov regularisation) was applied to neuroanatomical features to predict psychopathology factor scores. The objective function, a conjunction of a least squares loss and L^2 regularisation parametrised by coefficient λ , is given by:

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$$

We used a subset of $n = 8566$ unique participants from the dataset. We performed nested cross-validation with an outer loop on 15 stratified (on age and sex) folds and a 5-folds inner loop to fit the regularisation coefficient λ (alpha in `scikit-learn` implementation) in values $[1, 10, 100]$. Samples were normalised by total intracranial volume and then to unit norm.

3.3.2 Self-supervised contrastive learning

3.3.2.1 Architectures

To test our hypothesis that incorporating clinical variables in training would improve representation we compared two contrastive learning paradigms, see section 2.3.3 for details:

- **SimCLR** (Chen et al., 2020): A standard unsupervised framework using augmented image pairs to maximize agreement between representations of the same sample. It relies on images only and no additional labels.
- **yAware contrastive learning** (Dufumier et al., 2021): A weakly-supervised approach that incorporates auxiliary behavioural labels through a kernel into the contrastive loss, pushing negative pairs further away as their auxiliary labels are more different. We introduce three auxiliary metadata into our comparative study:

- **age-Aware**: the original version of y-Aware incorporating age as its auxiliary label proved to provide good results with transfer learning.
- **psy-Aware**: our model incorporating the longitudinal three-dimensions scale derived previously as the auxiliary labels. The metadata is thus multidimensional in this case.
- **neurodev-Aware**: the model incorporating a single factor (neurodevelopmental) from the previous three, we chose the one best predicted from VBM features (see Figure 4.2 in results).

3.3.2.2 Hyperparameters choice

yAware kernel Making a parallel with kernel density estimation (Dufumier, 2022), we estimated the kernel’s bandwidth Σ using Silverman’s rule of thumb: $\Sigma_{k,l} = 0$ if $k \neq l$, $\Sigma_{k,k} = \frac{n(d+2)^{-2/(d+4)}}{4} \sigma_k^2$ with $d = 3$ LP-dimensions and σ_k^2 the variance of the k -th factor.

Data preprocessing MRI volumes preprocessed with VBM were cropped and padded to a 128x128x128 shape, we ensured that no part of the brain was cropped in the process. We masked out the out-of-brain voxels (from MNI template) and z-normalised the values within the template.

Data augmentation To generate different views of MRI images, we combined random cropping with resizing, using the same parameters as in the seminal SimCLR study (Chen et al., 2020) and random erasing as proven effective in Dufumier et al. (2021) where we selected a sub-volume that is replaced with its mean value. We added light Gaussian noise to prevent the model from identifying positive pairs of augmented views based on background voxels values. We defined three strengths of data augmentation, defined by how much of the image can be cut out in random erasing: **light** with up to 20% of the volume erased, **medium** up to 40% and **strong** up to 60%. The exact parameters of data augmentation are found in Appendix A.3 and illustrations of the data augmentation are found in Figure A.1.

Optimiser hyperparameters Stochastic gradient descent was performed with AdamW (Loshchilov & Hutter, 2019) optimiser and a cosine decay scheduler (following guidelines from Balestriero et al. (2023)). We used grid search cross-validation to choose the initial learning rate in values

$[10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$ and the weight decay in values $[10^{-2}, 10^{-3}, 10^{-4}]$. We found 10^{-4} for the learning rate and 10^{-3} for the weight decay to be satisfactory values. We set the batch size to 32 samples per GPU and use 4 GPUs resulting in an effective batch size of 128 samples.

3.4 Downstream tasks and evaluation

We evaluated the representations learnt by contrastive learning on three downstream tasks. Evaluation was carried out via linear probing, meaning that a simple linear regressor or classifier was fitted on top of the model's embeddings without finetuning the model further on the downstream task.

3.4.1 Downstream tasks

We assessed model generalisability on three binary classification tasks (schizophrenia, bipolar disorder and autism spectrum disorder) using the out-of-sample clinical datasets described in section 3.2. We fitted a logistic regression classifier on top of the frozen embeddings for each model. Fitting was performed with nested cross-validation. The outer loop uses Monte Carlo Cross-validation (see next paragraph). The inner loop was used to fit the regularisation parameter (C in `scikit-learn` implementation) among values $[0.01, 0.1, 1, 10, 100]$ across 5 stratified folds. The classifier produced continuous prediction scores (in $[0, 1]$) for the positive class, from which the Receiver Operating Characteristic (ROC) curves were obtained by varying the decision threshold across all possible values. The Area Under the ROC Curve (AUROC) was then computed as a classification performance.

To get an accurate and generalisable estimate of the frozen models' performances on the small-scale downstream datasets, we used Monte Carlo cross-validation for evaluation following best practice (Varoquaux & Colliot, 2023). We generated 200 splits by stratified random sampling of MRI scans for the training (60% of dataset size) and testing (40% of dataset size) sets with replacement and ensuring that a same subject (or siblings) is not present in both sets. Also some of the data has been used for validation of the hyperparameters (data augmentation), we use this data in the training sets only to ensure no data leakage.

3.4.2 Statistical testing

For the prediction of LP-dimensions using Ridge regression, we performed a permutation test (100 permutations) by shuffling the dimensions among participants and repeating the nested cross-validation procedure on the obtained data.

To assess statistical significance in each downstream task, we applied the corrected resampled t -test proposed by Nadeau and Bengio (2003). This method adjusts the variance estimator by a factor of $(\frac{1}{J} + \frac{n_{test}}{n_{train}})$ to account for the correlation between test errors introduced by the overlap of training sets in Monte Carlo Cross-Validation (with $J = 200$ the number of cross-validation folds). We used it to test pairwise if one model outperformed the other and we used Bonferroni correction for multiple comparisons.

3.5 y-Aware loss visualisations

To better understand the behaviour of the psy-Aware model, we propose a novel visualisation of the y-Aware loss. As described in section 2.3.5, the y-Aware loss can be seen as the product of two terms: the SimCLR’s NT-XEnt loss and a modulating metadata similarity weight. The structure of the embedding space after training will be directly shaped by these weights, the original goal of the y-Aware framework being to organise the embedding space according to the auxiliary metadata. We propose to visualise the pairwise weights matrix for a batch as a heatmap with samples sorted according to the metadata of reference. For the psy-Aware model with three-dimensional metadata, we sort the samples according to the psychopathological burden defined as the norm of the vector of LP-dimensions.

From the weights matrix we can define a notion of neighbours for two samples i, k such that $weight(i, k)$ is greater than a predefined threshold (we set it to 0.01). We plot the distribution of the number of neighbours according to the metadata value and put it in regard to the distribution of the metadata.

To get insights into the structure of the embedding space, we also project the embedding space in 2D using UMAP (McInnes et al., 2020) and we colour the points by metadata values. Reproducing Dufumier et al. (2021) we expect a visual continuous colour gradient indicating that the embedding space is organised along the metadata of reference.

We take as a reference of desired behaviour the age-Aware model, previously described in Dufumier et al. (2021), that was pretrained on the large-scale UK Biobank cohort (“UK Biobank,” 2025) ($n = 42,786$

participants and scans) by a previous student at the lab.

3.6 Implementation

We ran model training on 4 Nvidia V100 GPUs with 32GB of memory and 160 GB of CPU memory on the Jean-Zay supercomputer. A typical training over 100 epochs took 40 GPU-hours.

All codes were written in Python 3.12. For exploratory factor analysis, we used the `factor_analyzer` (“factor_analyzer,” [n.d.](#)) package and `semopy` (Meshcheryakov et al., [2021](#)) for the implementation of polychoric correlations. Deep learning frameworks were implemented using `lightning` (Falcon & The PyTorch Lightning team, [2019](#)) and we used `scikit-learn` (Pedregosa et al., [2011](#)) for Ridge regression classifier. For 3D imaging data augmentation, we used the `torchio` (Pérez-García et al., [2021](#)) and `nidl` (“Nidl,” [n.d.](#)) libraries. Two months of the master’s thesis were dedicated to the development of the full training and evaluation pipeline distributed over several GPUs, based on the previously mentioned libraries. This work proved more complex than expected to produce a reliable and robust pipeline for distributed contrastive learning on 3D volumes. For the interested reader, the Python library `nidl` (“Nidl,” [n.d.](#)) now implements such training pipeline (and more) but it was not available at the start of this project.

Chapter 4

Results

4.1 CBCL-derived continuous scales of psychopathology

4.1.1 On baseline data, our analysis reproduces the results from Michelini et al.'s study

To derive a dimensional scale of psychopathology in youth, we first replicated the method from Michelini et al. (2019) and applied EFA on baseline CBCL data (see 3.1.3) with 1 to 5 extracted factors. We observe high mean absolute correlation between our factor solutions and those reported by Michelini et al. (2019). The mean Pearson's correlations over all factors are of $r = 0.98$ for the one-factor (p-factor) solution, $r = 0.99$ for two factors, $r = 0.94$ for three factors, $r = 0.97$ and $r = 0.98$ with four and five factors respectively. All p-values are under 10^{-27} per factor over 86 items and we used data from $n = 6,290$ subjects to derive the scales.

For the three-factor scale we name the dimensions *externalising* for the factor comprising items such as "Composite - Attacks/Threatens", "Argues a lot" or "Gets in many fights"; *internalising* for the factor with items "Worries", "Too fearful or anxious" or "Feels too guilty"; and *neurodevelopmental* for items like "Stares blankly", "Composite - Distracted/Hyperactive" or "Confused or seems to be in a fog", following Michelini et al. (2019)'s nomenclature. We focus solely on this scale for the subsequent analyses.

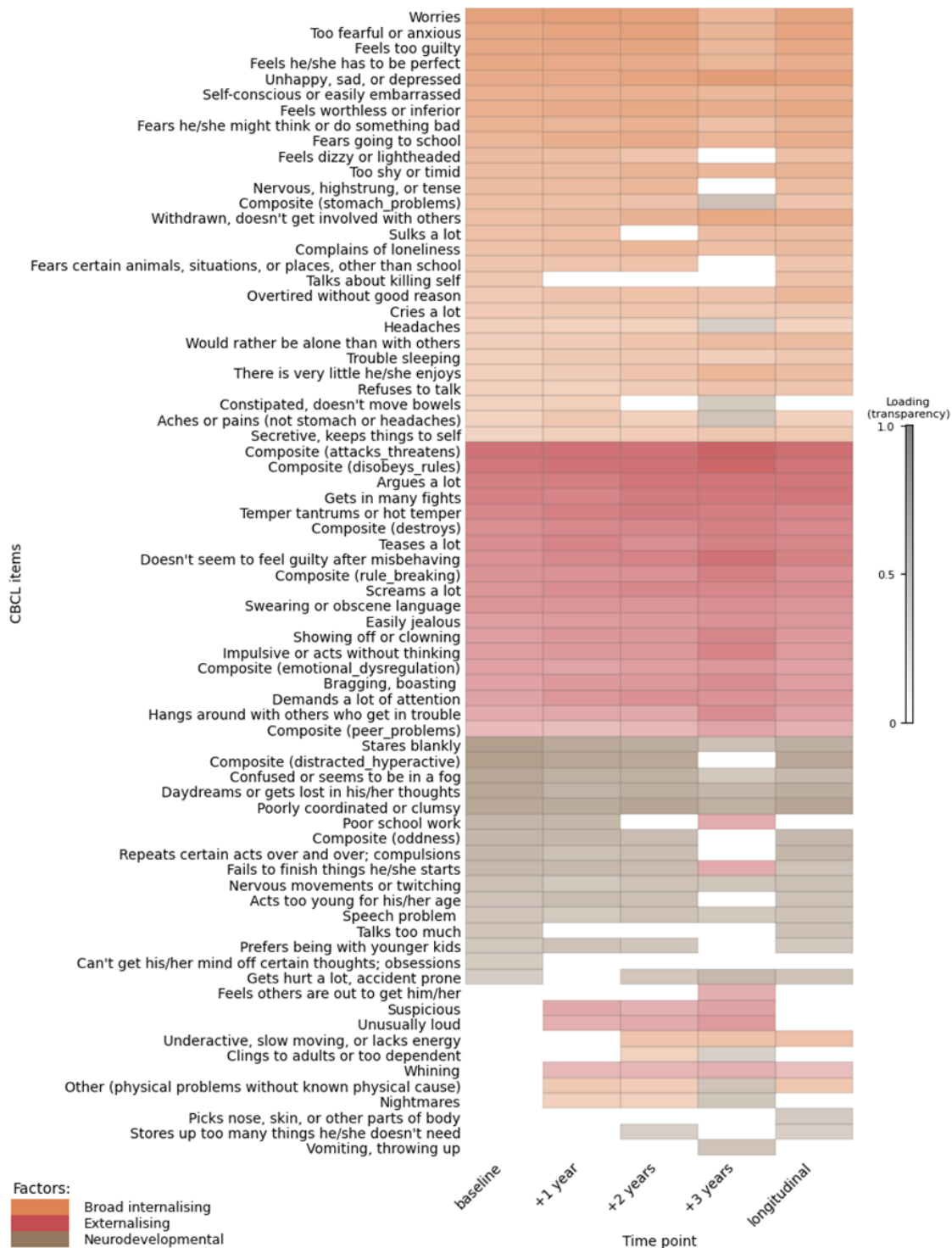


Figure 4.1: **EFA loadings across time points** — Each panel represents the loading matrix for the EFA three-factor solution using baseline and successive follow-ups data. The longitudinal model uses data from all time points. Cell colour indicates the factor on which an item primarily loads, and transparency reflects the magnitude of the loading coefficient. CBCL items that load primarily on a factor in the baseline model are ordered from highest to lowest loading within each factor. After them are the factors that load for other models but not for baseline. Items that do not load primarily on a factor in any of the scales are not displayed.

4.1.2 The three-factor structure is stable across time

In order to identify longitudinal factors in youth, we then extended the work of Michelini et al. (2019)'s on the three factors scale by deriving factors of psychopathology using independently data from the follow-up data collections in the ABCD study (at +1, +2 and +3 years) and with data from all time points (longitudinal). We report correlation of each factor with respect to baseline in Table 4.1. We expect high correlation in the first years as the clinical symptoms of participants should not, overall, change abruptly from one year to the other. We observe high stability for all factors and time points ($r > 0.9$) except for the neurodevelopmental dimension at +3 years ($r = 0.67$). Visual inspection of the loading matrices (Figure 4.1) further supports the quantitative results: we observe consistency across time points but less distinct in the +3 years neurodevelopmental dimension. Surprisingly in this case, the composite item "Distracted/Hyperactive" does not load primarily on the neurodevelopmental factor while it is one of the highest loadings for the other time points. Actually, its loading on the neurodevelopmental factor is higher than the 0.35 threshold (loading of 0.47) but because it also loads highly on the externalising factor (loading of 0.46) we cannot assign it to an unambiguous single factor (see methods 3.1.3 for details). This observation may reflect the changes in psychopathology symptoms across development as +3 years is the furthest point from the baseline data. We also observe a small decrease in correlation for the other factors at +3 years. In the following analysis we use the longitudinal scale which loadings are reported in Appendix B.1.

Table 4.1: **Stability of loadings across time points** — Correlations between baseline and follow-up factor loadings for the three-factor solution. The longitudinal solution is derived using data from all time points. Rows represent factors named after the terminology of Michelini et al. (2019); columns represent time points.

Factor name	Baseline	+1 year	+2 years	+3 years	Longitudinal
Broad internalising	1.00	0.98	0.98	0.91	0.96
Externalising	1.00	0.99	0.99	0.96	0.99
Neurodevelopmental	1.00	0.97	0.97	0.67	0.94

4.2 Externalising and neurodevelopmental dimensions are better predicted from VBM than the internalising symptoms

To ensure there is a minimal correlation between brain anatomy and the chosen behavioural dimensions, we fitted linear ridge regression models to predict the latent psychopathology dimensions from imaging features (VBM preprocessing) at the region level. The correlation between predicted and true values of the dimensions reaches about 10% (Figure 4.2) for the externalising and developmental dimensions indicating a challenging task but aligned with previous results in the literature (Ooi et al., 2022) and performing better than chance (p-value = 0.01). The internalising factor however is predicted at chance level suggesting it may have lower or more complex correlation to neuroanatomy. Overall these results encourage to continue with the planned approach and include psychopathology dimensions into the weakly supervised deep learning framework.

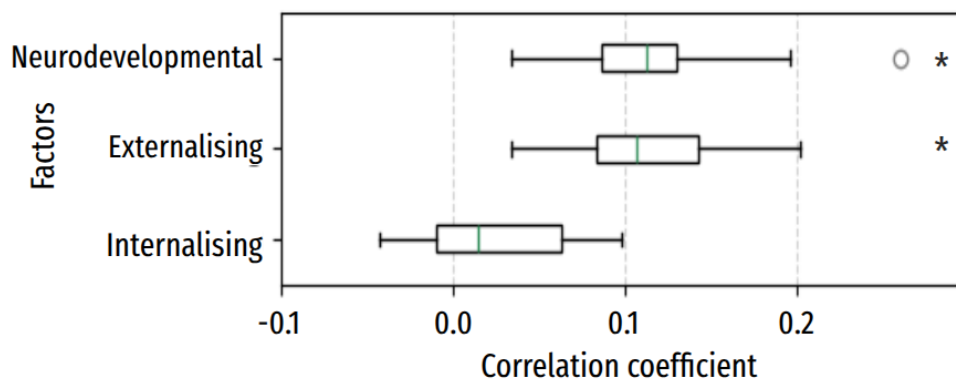


Figure 4.2: **Linear prediction of latent psychopathology dimensions from ROI features** — Boxplots of correlation between the true dimensions and the values predicted by a ridge regressor over 15 stratified folds on n=8,566 subjects. * indicates p-value<0.01. Input is ROI features of grey matter volumes defined on the neuromorphometrics atlas.

4.3 Incorporating dimensions of psychopathology yields comparable performances to standard contrastive learning

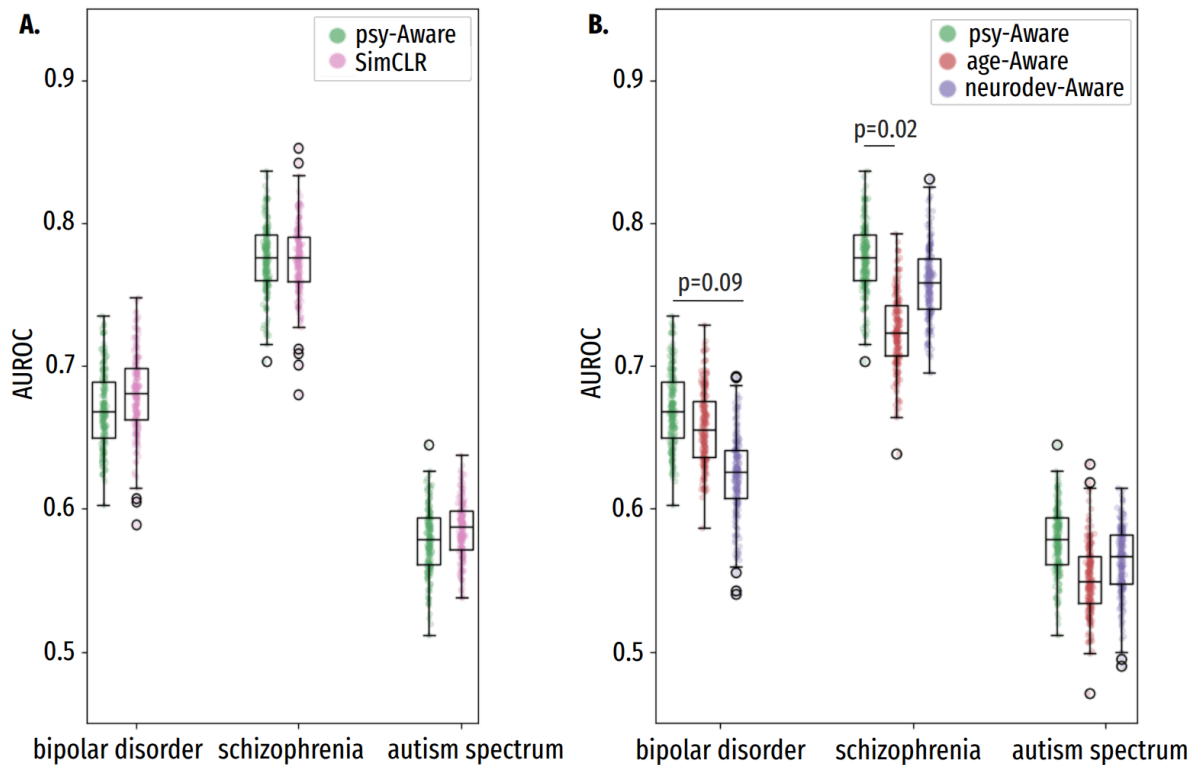


Figure 4.3: **Downstream diagnosis classification results** — Performance on downstream tasks measured with area under the ROC curve (AUROC) and evaluated over 200 folds of Monte Carlo cross-validation. P-values under 0.1 after Bonferroni correction are indicated. **A.** Comparison of weakly supervised contrastive learning model psy-Aware and baseline contrastive learning model SimCLR. **B.** Comparison for different auxiliary variables incorporated in y-Aware: all latent psychopathology dimensions (psy-Aware), age (age-Aware) and the dimension best predicted from VBM (neurodev-Aware).

To evaluate our method, psy-Aware, against SimCLR, we fit a simple linear probe on top of the embeddings resulting from each model and compare their performances on three diagnosis downstream tasks. Incorporating psychopathology dimensions with the psy-Aware model yields equivalent performance to standard contrastive learning in predicting psychiatric

diagnosis, achieving comparable AUROC values, see Figure 4.3A.

To better evaluate the influence of the auxiliary metadata choice in *y*-Aware, we also compared *psy*-Aware to other *y*-Aware derived models: the robust baseline *age*-Aware, incorporating age in the kernel, and to *neurodev*-Aware, a model incorporating a single factor of psychopathology (neurodevelopmental factor which is predicted better than chance from VBM features, Figure 4.2). *Psy*-Aware, incorporating all dimensions, achieves higher AUROC than *neurodev*-Aware incorporating a single factor in the bipolar disorder classification task (p-value=0.09, Figure 4.3B). *Psy*-Aware performs better than *age*-Aware on the schizophrenia classification, (p-value=0.02, Figure 4.3B).

4.3.1 Hyperparameters search

For SimCLR, *psy*-Aware and *age*-Aware, we cross-validated the data augmentation regime on the training and validation datasets, results are reported in Table 4.2. Overall the variation in performance was low for the weakly supervised models, they had the highest estimated mean value with the lightest data augmentation so we chose this hyperparameter for the experiments though we admit that this choice is somewhat arbitrary. We extended this choice to train the *neurodev*-Aware model. For SimCLR, stronger data augmentation led to better performances so we used the strongest data augmentation.

Table 4.2: **Hyperparameters search for data augmentation** — Mean AUROC results on all three downstream tasks (SCZ, BD, ASD) over 100 folds of Monte Carlo cross-validation on the validation dataset. In bold is the best point-estimate performance for each model. Light, Medium and Strong refers to the data augmentation strength.

Model	Light	Medium	Strong
SimCLR	0.630	0.639	0.641
<i>psy</i> -Aware	0.649	0.637	0.649
<i>age</i> -Aware	0.642	0.634	0.636

4.4 The highly skewed factor distribution may deteriorate training quality

To understand further why the weak supervision seems to have little influence on the structure of the embedding space, we visualise the weights matrix from the psy-Aware loss for a representative batch of size 128, the batch size used during training (Figure 4.5A). Approximately 20 samples share identical factor values (0,0,0), resulting in uniform weights between them. Conversely, samples with higher psychopathological burden share strong weights with only a few other samples. This creates a skewed weight distribution. In contrast, the age-Aware model trained on UKB shows a stable weight distribution, with a well-defined diagonal of high values across the batch (Figure 4.5B).

The downstream performances of psy-Aware are not informative on the learnt features by the model. To assess whether the embedding space is structured according to the LP-dimensions, we visualise the 2D UMAP projection of the embedding space (see Figure 4.4A). We would expect a continuous colour gradient such as in the reference y-Aware model (age-Aware trained on the UKB cohort, Figure 4.4B) but we observe no visible structure of the projection along the LP-dimensions.

Extending this analysis across a full epoch (133 batches covering the ABCD dataset), we observe that low-burden samples have many neighbours, up to 40 per batch (representing 30% of the batch's samples), while high-burden samples have very few (≤ 5 , sometimes none; Figure 4.6B). This pattern mirrors the skewed distribution of individual dimensions (Figure 4.6C).

The skewed neighbour distribution (Figure 4.6B) reduces resolution among low-burden subjects, effectively collapsing them into a single categorical label (“no psychopathology”), which produces a diffuse and ambiguous learning signal. Conversely, high-burden subjects receive minimal supervision, leaving the algorithm to behave nearly as unsupervised SimCLR. These two edge cases, both present in each batch, may explain why the embedding space fails to structure according to LP-dimensions (as observed in UMAP projection, Figure 4.4A).

The proposed psy-Aware framework assumes that metadata varies continuously across individuals. However, we observed that approximately 30% of the cohort exhibited no CBCL symptoms and therefore shared identical psychopathology factor scores (0,0,0). This ceiling effect substantially reduces phenotypic variance and limits the effectiveness of dimensional learning approaches.

We also note that there are domain gaps between ABCD, the training cohort, and the clinical datasets used for downstream tasks. Differences in age distributions (ABCD (mean age: 11 years) versus clinical cohorts (mean ages: 36 for BD, 33 for SCZ, and 16 for ASD)) as well as variations in acquisition protocols, are likely to affect the transferability of learnt representations. These discrepancies may partially explain the reduced performance observed in clinical applications.

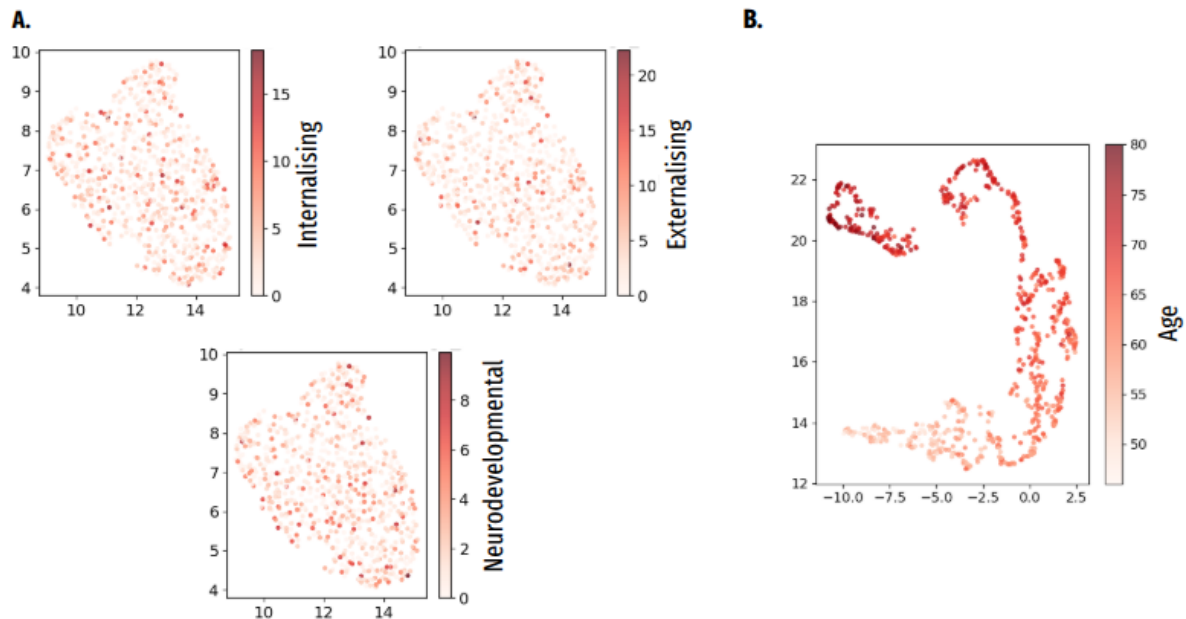


Figure 4.4: **UMAP projections of embedding space** — 2D UMAP projection of 800 training samples embeddings from **A.** Psy-Aware trained on ABCD with points coloured according to the value of each factor of psychopathology, **B.** Age-Aware trained on UKB with points coloured according to age.

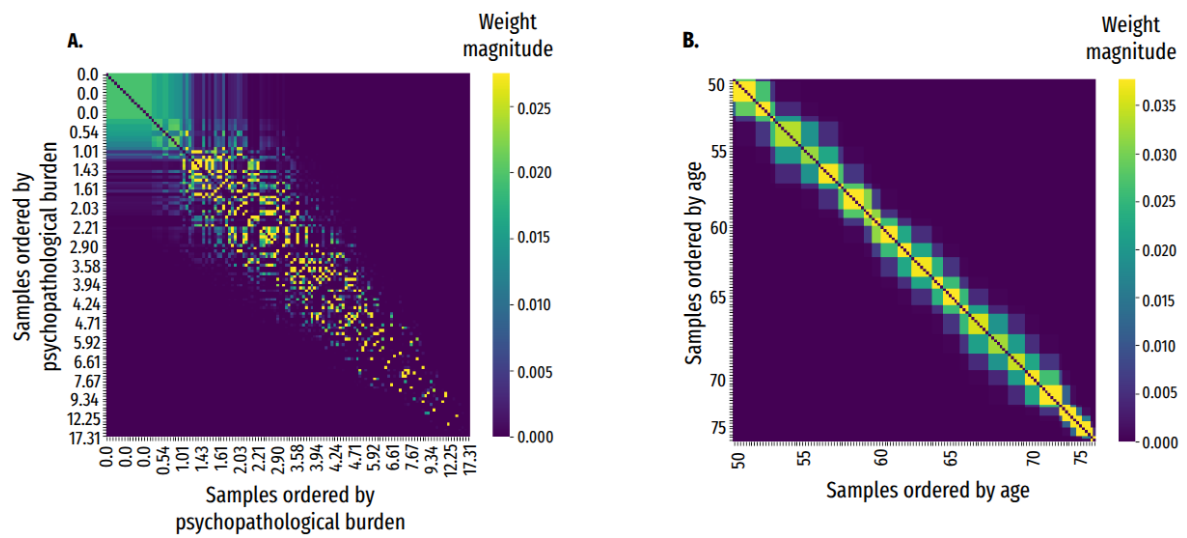


Figure 4.5: **Weights matrix from y-Aware loss** — Weights matrix for a training batch of 128 samples ordered along the metadata of references. Colour represents the weights magnitudes. **A.** Psy-Aware model trained on ABCD with three latent psychopathology dimensions as metadata, psychopathological burden is defined as the norm of the metadata vector. **B.** Age-Aware model trained on UKB with age as the metadata of reference.

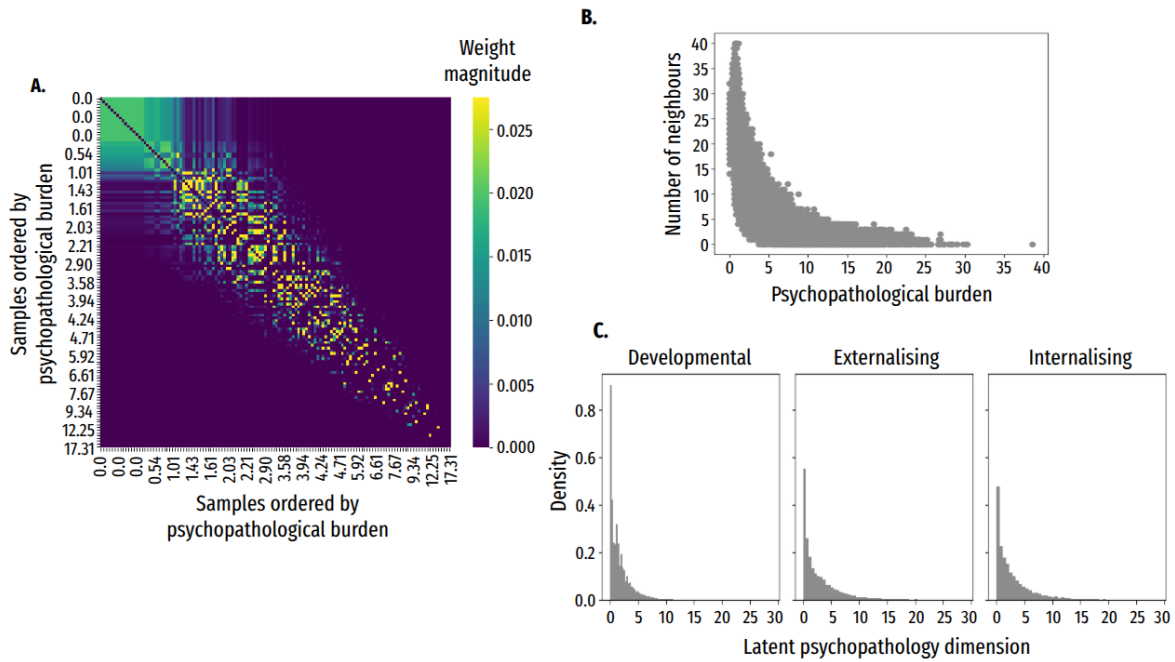


Figure 4.6: **Parallel between distribution of neighbours in a batch and distribution of metadata** — **A.** In a batch of size 128, the weight matrix of psy-Aware loss is unbalanced as samples with low psychopathological burden have more neighbours than those with a high psychopathological burden. **B.** Distribution of the number of neighbours in a batch according to the psychopathological burden across 133 batches (full training set). Each point represents one sample in one batch. **C.** Latent psychopathology dimensions' histograms.

Chapter 5

Discussion and limitations

This study investigates whether incorporating latent psychopathology dimensions (LP-dimensions) into weakly supervised contrastive learning can improve the clinical relevance of brain anatomy representations. We first derived a longitudinal three dimensions model of youth psychopathology using factor analysis, effectively reproducing then extending the study by Michelini et al. (2019).

To ensure a minimal brain-behaviour correlation, we then predicted these LP-dimensions from regional volumes of grey matter and obtained limited but state-of-the art accuracy aligned with previous work (Ooi et al., 2022).

Last, we compared our weakly-supervised proposed approach, psy-Aware, to the fully unsupervised framework SimCLR on three diagnosis tasks in psychiatry and observe that the models perform similarly. While both models' hyperparameters were reasonably optimised, we cannot rule out that different hyperparameters could yield better downstream performances. Additionally, we have chosen two approaches based on contrastive learning, in the aim of studying the particular effect of incorporating LP-dimensions into training. However other non-contrastive self-supervised methods such as DINO (Caron et al., 2021) could represent interesting alternative frameworks to learn neuroanatomy representations.

Apart from this general discussion, we identified several more specific limitations to our work described in the next sections.

5.1 Factor analysis is restricted to young teenagers

Factor analysis is limited by the age range of the ABCD dataset in release 5 (9 to 15 years old). We excluded items endorsed in less than 1% of the sample but those items are mostly problems arising in later teenage years (e.g.: alcohol and drugs consumption, sex problems...) therefore making the longitudinal latent psychopathology dimensions (LP-dimensions) derived with EFA not so applicable to older teenagers. Leveraging the subsequent releases of the ABCD data could alleviate this issue.

5.2 Skewed metadata distribution limits the application of y-Aware

We observed that the LP-dimensions have a low variance in the dataset used for pretraining. We hypothesise that y-Aware requires greater variability in the metadata to meaningfully structure the embedding space. This is supported by the poor performance of age-Aware on ABCD (Figure 4.3B), a youth dataset with limited age range (9-15 years old), in contrast to its strong performance when pretrained on datasets spanning broader age ranges (Dufumier et al., 2024).

5.3 Low resolution of latent psychopathology dimensions in healthy subjects

The psychopathology scale derived from the CBCL items has low resolution for healthy subjects which limits the effectiveness of the dimensional approach. Recent works have raised concerns regarding the noise in mental health measures (Schulz et al., 2024), and the replicability and clinical interpretability of factor structures derived from CBCL-like instruments (Tiego et al., 2023). As the CBCL was originally designed to distinguish referred from non-referred children rather than to characterise the full dimensional spectrum of psychopathology (Achenbach, 2001), its derived LP-dimensions offer limited sensitivity in low-symptom populations and are susceptible to method variance and informant effects (Tiego et al., 2023). One potential mitigation strategy is to integrate multiple complementary instruments from

different informants (youth, parents, teachers), thereby increasing phenotypic resolution and improving the robustness of dimensional assessments.

Chapter 6

Conclusions and Future work

6.1 Conclusions

This thesis integrated several innovative approaches to address a longstanding challenge in psychiatry: linking neuroimaging to psychiatric phenotypes. Our purpose was to generate low-dimensional representations of brain anatomy that are predictive of psychiatric diagnosis. We investigated whether incorporating continuous latent psychopathology dimensions (LP-dimensions) into weakly supervised contrastive learning using the y -Aware loss and the large-scale ABCD dataset could yield such representations and improve the clinical relevance over standard contrastive learning. We first derived and validated a longitudinal three LP-dimensions structure (externalising, internalising and neurodevelopmental) in youth, confirming its stability across early adolescence. When incorporated into the psy-Aware framework, these LP-dimensions yielded embeddings with comparable clinical downstream performance to standard contrastive learning. Importantly, we identified that the skewed distribution of LP-dimensions in ABCD, particularly the prevalence of low-burden participants, limits the ability of y -Aware to structure embeddings meaningfully, partially constraining the goal of generating clinically informative representations. Overall, this work highlights the critical role of metadata variance for weakly supervised training and the limitations of widely used instruments like the CBCL for dimensional modelling, providing valuable insights for future studies at the interface of neuroimaging and psychiatry.

6.2 Future works

Dimensional psychiatry assessment tool We've only considered one tool for behavioural measures (the CBCL) to produce a dimensional scale of psychopathology. Future works could focus on deriving dimensional psychiatric scales with a higher resolution, for instance by incorporating measures from different informants (youth, parents, teachers) or from different instruments.

Adaptive y-Aware kernel's bandwidth Inspired by the Stochastic Neighbour Embedding (SNE) algorithm (Hinton & Roweis, 2002), future works could use an adaptive kernel bandwidth Σ in the y-Aware loss such that each sample has approximately the same number of neighbours (as a parallel to the concept of perplexity in SNE).

Imaging preprocessing We used VBM preprocessing which comprises some heavy normalising steps (e.g.: it projects all images onto the same template). While this helps in reducing site effect it may alter important inter-individual variability in neuroanatomy. Moreover, we only focused on volumes of grey matter but other neuroanatomical features of interest could be investigated through with lighter imaging preprocessing preserving more of the original signal. Future works could work on minimally processed images, like the quasi-raw preprocessing describes in Dufumier et al. (2024).

Data augmentation Future works could explore advanced data augmentation strategies to mitigate the domain gaps between the pretraining and downstream data to improve generalisation.

Multimodality Our work used T1w MRI imaging but other MRI contrasts such as T2w and FLAIR can reveal different informations of brain anatomy. A future direction could be to extend the present study to other MRI modalities or to fuse them into a multimodal model.

Interpretability Similar downstream results between y-Aware and SimCLR are not informative of the features captured by one or the other model. We've used visualisations to get further insights into the organisation of the embedding space but those are still shallow analysis. Future work could investigate the differences in the resulting embeddings spaces and bring more interpretability into the downstream tasks results.

6.3 Reflections

This thesis contributes to research in medical imaging and to our understanding of the neural basis of psychiatric disorders. It aligns with the United Nations Sustainable Development Goal number 3 "Ensure healthy lives and promote well-being for all at all ages" and in particular to target 3.4 aiming to "...promote mental health and well-being".

In this work we've used human personal health data (medical imaging, clinical diagnoses) which pose specific risks to personal privacy. We've ensured to store and use data on adequate secured environments.

The boom of deep learning applications has a disastrous impact on the environment including, but not limited to, high carbon emissions, gigantic energy consumption and increased water usage and pollution (IEA, 2025; Li et al., 2025). To limit the footprint of this work, we've ensured to use a moderate size model (33M parameters) adapted to our tasks of interest rather than choosing deeper and larger models that are also less energy-efficient (Varoquaux et al., 2025). We've tracked the carbon emissions of our models' training using the CEEMS tool ("ceems-dev/ceems," 2026) available on Jean Zay's computing nodes. It amounts to 21.4 kg CO₂-equivalent for the whole thesis, which compares to roughly 100 km using a thermal car*. The Jean-Zay supercomputer being hosted in France, it benefits from the French low-carbone electricity mix (IEA, n.d.).

We also acknowledge that deep learning models in medical imaging often exhibit biases based on sex or ethnicity. In this study we normalised the grey matter volumes in VBM images by the intracranial total volume, which reduces bias across sex and ethnicity (Liu et al., 2025). Although due to time constraints we haven't systematically monitored possible biases in the models' embeddings, it should be monitored and mitigated in future works.

Large language models were used for edition of the manuscript and code prototyping. The author proofread all generated text and scripts and takes full responsibility for the work.

*Using the methodology from the The French Agency for Ecological Transition at <https://impactco2.fr/outils/comparateur>.

References

- ABCD study* [ABCD study]. (n.d.). Retrieved July 4, 2025, from <https://abcdstudy.org/>
- ABIDE*. (n.d.). Retrieved November 5, 2025, from https://fcon_1000.projects.nitrc.org/indi/abide/
- Abrol, A., Bhattarai, M., Fedorov, A., Du, Y., Plis, S., & Calhoun, V. (2020). Deep residual learning for neuroimaging: An application to predict progression to alzheimer's disease. *Journal of Neuroscience Methods*, 339, 108701. <https://doi.org/10.1016/j.jneumeth.2020.108701>
- Achenbach, T. M. (2001). *Manual for the ASEBA school-age forms & profiles : An integrated system of multi-informant assessment*. Burlington, VT : ASEBA. Retrieved July 15, 2025, from <http://archive.org/details/manualforasebasc0000ache>
- Achenbach, T. M. (2020). Bottom-up and top-down paradigms for psychopathology: A half-century odyssey [Publisher: Annual Reviews]. *Annual Review of Clinical Psychology*, 16, 1–24. <https://doi.org/10.1146/annurev-clinpsy-071119-115831>
- Achenbach, T. M., ... Verhulst, F. C. (2025). P-factor(s) for youth psychopathology across informants and models in 24 societies [Publisher: Routledge _eprint: <https://doi.org/10.1080/15374416.2024.2344159>]. *Journal of Clinical Child & Adolescent Psychology*, 54(3), 318–327. <https://doi.org/10.1080/15374416.2024.2344159>
- Aglinskas, A., Hartshorne, J. K., & Anzellotti, S. (2022). Contrastive machine learning reveals the structure of neuroanatomical variation within autism [Publisher: American Association for the Advancement of Science]. *Science*, 376(6597), 1070–1074. <https://doi.org/10.1126/science.abm2461>
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1), 95–113. <https://doi.org/10.1016/j.neuroimage.2007.07.007>

- Balestriero, R., Ibrahim, M., Sobal, V., Morcos, A., Shekhar, S., Goldstein, T., ... Goldblum, M. (2023, June 28). A cookbook of self-supervised learning [Issue: arXiv:2304.12210]. <https://doi.org/10.48550/arXiv.2304.12210>
- Bashyam, V. M., Erus, G., Doshi, J., Habes, M., Nasrallah, I. M., Truelove-Hill, M., ... Davatzikos, C. (2020). MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain*, *143*(7), 2312–2324. <https://doi.org/10.1093/brain/awaa160>
- Billot, B., Magdamo, C., Cheng, Y., Arnold, S. E., Das, S., & Iglesias, J. E. (2023). Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *120*(9), e2216399120. <https://doi.org/10.1073/pnas.2216399120>
- Blok, E., Geenjaar, E. P. T., Geenjaar, E. A. W., Calhoun, V. D., & White, T. (2022). Neurodevelopmental trajectories in children with internalizing, externalizing and emotion dysregulation symptoms. *Frontiers in Psychiatry*, *13*, 846201. <https://doi.org/10.3389/fpsy.2022.846201>
- Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., ... Gur, R. E. (2015). The philadelphia neurodevelopmental cohort: Constructing a deep phenotyping collaborative. *Journal of child psychology and psychiatry, and allied disciplines*, *56*(12), 1356–1369. <https://doi.org/10.1111/jcpp.12416>
- CANDI. (n.d.). Retrieved November 5, 2025, from https://www.nitrc.org/projects/candi_share
- Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers, 9650–9660. Retrieved July 4, 2025, from https://openaccess.thecvf.com/content/ICCV2021/html/Caron_Emerging_Properties_in_Self-Supervised_Vision_Transformers_ICCV_2021_paper
- Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*(1), 41–75. <https://doi.org/10.1023/A:1007379606734>

- Ceems-dev/ceems* [original-date: 2023-11-15T10:40:30Z]. (2026, January 5). CEEMS Project. Retrieved January 6, 2026, from <https://github.com/ceems-dev/ceems>
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, June 30). A simple framework for contrastive learning of visual representations [Issue: arXiv:2002.05709]. <https://doi.org/10.48550/arXiv.2002.05709>
- Chopra, S., Dhamala, E., Lawhead, C., Ricard, J. A., Orchard, E. R., An, L., ... Holmes, A. J. (2024). Generalizable and replicable brain-based predictions of cognitive functioning across common psychiatric illness [Publisher: American Association for the Advancement of Science]. *Science Advances*, *10*(45), eadn1862. <https://doi.org/10.1126/sciadv.adn1862>
- Costafreda, S. G., Chu, C., Ashburner, J., & Fu, C. H. Y. (2009). Prognostic and diagnostic potential of the structural neuroanatomy of depression [Publisher: Public Library of Science]. *PLOS ONE*, *4*(7), e6353. <https://doi.org/10.1371/journal.pone.0006353>
- Delvecchio, G., Fossati, P., Boyer, P., Brambilla, P., Falkai, P., Gruber, O., ... Frangou, S. (2012). Common and distinct neural correlates of emotional processing in bipolar disorder and major depressive disorder: A voxel-based meta-analysis of functional magnetic resonance imaging studies. *European Neuropsychopharmacology*, *22*(2), 100–113. <https://doi.org/10.1016/j.euroneuro.2011.07.003>
- Dhamala, E., Yeo, B. T. T., & Holmes, A. J. (2023). One size does not fit all: Methodological considerations for brain-based predictive modeling in psychiatry. *Biological Psychiatry*, *93*(8), 717–728. <https://doi.org/10.1016/j.biopsych.2022.09.024>
- Dufumier, B. (2022, December 16). *Representation learning in neuroimaging : Transferring from big healthy data to small clinical cohorts* (These de doctorat). université Paris-Saclay. Retrieved November 5, 2025, from <https://theses.fr/2022UPASG093>
- Dufumier, B., Gori, P., Petiton, S., Louiset, R., Mangin, J.-F., Grigis, A., & Duchesnay, E. (2024). Exploring the potential of representation and transfer learning for anatomical neuroimaging: Application to psychiatry. *NeuroImage*, *296*, 120665. <https://doi.org/10.1016/j.neuroimage.2024.120665>
- Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., ... Duchesnay, E. (2021). Contrastive learning with continuous proxy meta-data for 3d MRI classification. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International*

- Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II*, 58–68. https://doi.org/10.1007/978-3-030-87196-3_6
- EducationalTestingService/factor_analyzer: A python module to perform exploratory & confirmatory factor analyses*. (n.d.). Retrieved August 6, 2025, from https://github.com/EducationalTestingService/factor_analyzer
- Falcon, W., & The PyTorch Lightning team. (2019, March). *PyTorch lightning* (Version 1.4). <https://doi.org/10.5281/zenodo.3828935>
- Fernandes, B. S., Williams, L. M., Steiner, J., Leboyer, M., Carvalho, A. F., & Berk, M. (2017). The new field of ‘precision psychiatry’. *BMC Medicine*, *15*(1). <https://doi.org/10.1186/s12916-017-0849-x>
- Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., Luders, E., & the Alzheimer’s Disease Neuroimaging Initiative. (2024). CAT: A computational anatomy toolbox for the analysis of structural MRI data. *GigaScience*, *13*, giae049. <https://doi.org/10.1093/gigascience/giae049>
- Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., ... Etkin, A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*, *72*(4), 305–315. <https://doi.org/10.1001/jamapsychiatry.2014.2206>
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, *40*(7), 3510–3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Gorgolewski, K. J., Durnez, J., & Poldrack, R. A. (2017). Preprocessed consortium for neuropsychiatric phenomics dataset. *F1000Research*, *6*, 1262. <https://doi.org/10.12688/f1000research.11964.2>
- Grønneberg, S., & Foldnes, N. (2024). Factor analyzing ordinal items requires substantive knowledge of response marginals. *Psychological Methods*, *29*(1), 65–87. <https://doi.org/10.1037/met0000495>
- Gutmann, M., & Hyvärinen, A. (2010). Noise-contrastive estimation: A new estimation principle for unnormalized statistical models [ISSN: 1938-7228]. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 297–304. Retrieved August 19, 2025, from <https://proceedings.mlr.press/v9/gutmann10a.html>
- Hill, E. D., Kashyap, P., Raffanella, E., Wang, Y., Moffitt, T. E., Caspi, A., ... Posner, J. (2025). Prediction of mental health risk in adolescents

- [Publisher: Nature Publishing Group]. *Nature Medicine*, 1–7. <https://doi.org/10.1038/s41591-025-03560-7>
- Hinton, G. E., & Roweis, S. (2002). Stochastic neighbor embedding. *Advances in Neural Information Processing Systems*, 15. Retrieved January 8, 2026, from https://proceedings.neurips.cc/paper_files/paper/2002/hash/6150ccc6069bea6b5716254057a194ef-Abstract.html
- Hoffmann, M. S., Moore, T. M., Axelrud, L. K., Tottenham, N., Rohde, L. A., Milham, M. P., ... Salum, G. A. (2023). Harmonizing bifactor models of psychopathology between distinct assessment instruments: Reliability, measurement invariance, and authenticity. *International Journal of Methods in Psychiatric Research*, 32(3), e1959. <https://doi.org/10.1002/mpr.1959>
- Hoffmann, M. S., Moore, T. M., Axelrud, L. K., Tottenham, N., Zuo, X.-N., Rohde, L. A., ... Salum, G. A. (2022). Reliability and validity of bifactor models of dimensional psychopathology in youth. *Journal of psychopathology and clinical science*, 131(4), 407–421. <https://doi.org/10.1037/abn0000749>
- IEA. (n.d.). *France - countries & regions* [IEA]. Retrieved January 8, 2026, from <https://www.iea.org/countries/france>
- IEA. (2025, April 10). *Energy and AI* [IEA]. Retrieved January 8, 2026, from <https://www.iea.org/reports/energy-and-ai>
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders [Publisher: American Psychiatric Publishing]. *American Journal of Psychiatry*, 167(7), 748–751. <https://doi.org/10.1176/appi.ajp.2010.09091379>
- Kaczurkin, A. N., Park, S. S., Sotiras, A., Moore, T. M., Calkins, M. E., Cieslak, M., ... Satterthwaite, T. D. (2019). Evidence for dissociable linkage of dimensions of psychopathology to brain structure in youths [Publisher: American Psychiatric Publishing]. *American Journal of Psychiatry*, 176(12), 1000–1009. <https://doi.org/10.1176/appi.ajp.2019.18070835>
- Kaczmarek, E., Szeto, J., Nichyporuk, B., & Arbel, T. (2025). Building a general SimCLR self-supervised foundation model across neurological diseases to advance 3d brain MRI diagnoses, 1310–1319. Retrieved January 8, 2026, from https://openaccess.thecvf.com/content/ICCV2025W/CVAMD/html/Kaczmarek_Building_a_General_SimCLR_Self-Supervised_Foundation_Model_Across_Neurological_Diseases_ICCVW_2025_paper.html

- Kandel, E. R. (1999). Biology and the future of psychoanalysis: A new intellectual framework for psychiatry revisited [Publisher: American Psychiatric Publishing]. *American Journal of Psychiatry*, 156(4), 505–524. <https://doi.org/10.1176/ajp.156.4.505>
- Kirschner, M., Shafiei, G., Markello, R. D., Makowski, C., Talpalaru, A., Hodzic-Santor, B., ... Mišić, B. (2020). Latent clinical-anatomical dimensions of schizophrenia. *Schizophrenia Bulletin*, 46(6), 1426–1438. <https://doi.org/10.1093/schbul/sbaa097>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., ... Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies [Place: US Publisher: American Psychological Association]. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). Making AI less 'thirsty'. *Commun. ACM*, 68(7), 54–61. <https://doi.org/10.1145/3724499>
- Liu, P., Zemlyanker, D., Gopinath, K., Cheng, Y., He, Y., Izquierdo-Garcia, D., ... Iglesias, J. E. (2025). The normalizing properties of intracranial volume across race and sex. *Brain Communications*, 7(4), fcaf271. <https://doi.org/10.1093/braincomms/fcaf271>
- Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution [Publisher: Routledge _eprint: <https://doi.org/10.1080/10705511.2020.1735393>]. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(1), 138–147. <https://doi.org/10.1080/10705511.2020.1735393>
- Loshchilov, I., & Hutter, F. (2019, January 4). Decoupled weight decay regularization. <https://doi.org/10.48550/arXiv.1711.05101>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals [Publisher: Nature Publishing Group]. *Nature*, 603(7902), 654–660. <https://doi.org/10.1038/s41586-022-04492-9>
- McInnes, L., Healy, J., & Melville, J. (2020, September 18). UMAP: Uniform manifold approximation and projection for dimension reduction. <https://doi.org/10.48550/arXiv.1802.03426>
- Meshcheryakov, G., Igolkina, A. A., & Samsonova, M. G. (2021, June 9). Semopy 2: A structural equation modeling package with random effects in python. <https://doi.org/10.48550/arXiv.2106.01140>

- Mewton, L., Lees, B., Squeglia, L. M., Forbes, M. K., Sunderland, M., Krueger, R., ... Teesson, M. (2022). The relationship between brain structure and general psychopathology in preadolescents. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 63(7), 734–744. <https://doi.org/10.1111/jcpp.13513>
- Michelini, G., Barch, D. M., Tian, Y., Watson, D., Klein, D. N., & Kotov, R. (2019). Delineating and validating higher-order dimensions of psychopathology in the adolescent brain cognitive development (ABCD) study [Number: 1]. *Translational Psychiatry*, 9(1), 261. <https://doi.org/10.1038/s41398-019-0593-4>
- Moore, T. M., Kaczkurkin, A. N., Durham, E. L., Jeong, H. J., McDowell, M. G., Dupont, R. M., ... Lahey, B. B. (2020). Criterion validity and relationships between alternative hierarchical dimensional models of general and specific psychopathology. *Journal of Abnormal Psychology*, 129(7), 677–688. <https://doi.org/10.1037/abn0000601>
- Nadeau, C., & Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, 52(3), 239–281. <https://doi.org/10.1023/a:1024068626366>
- Nidl [Nidl]. (n.d.). Retrieved January 8, 2026, from <https://neurospin-deepinsight.github.io/index.html>
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), 443–460. <https://doi.org/10.1007/BF02296207>
- Ooi, L. Q. R., Chen, J., Zhang, S., Kong, R., Tam, A., Li, J., ... Yeo, B. T. T. (2022). Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *NeuroImage*, 263, 119636. <https://doi.org/10.1016/j.neuroimage.2022.119636>
- Oord, A. v. d., Li, Y., & Vinyals, O. (2019, January 22). Representation learning with contrastive predictive coding. <https://doi.org/10.48550/arXiv.1807.03748>
- Ortiz-Gil, J., Pomarol-Clotet, E., Salvador, R., Canales-Rodríguez, E. J., Sarró, S., Gomar, J. J., ... McKenna, P. J. (2011). Neural correlates of cognitive impairment in schizophrenia. *The British Journal of Psychiatry: The Journal of Mental Science*, 199(3), 202–210. <https://doi.org/10.1192/bjp.bp.110.083600>
- Parkes, L., Moore, T. M., Calkins, M. E., Cook, P. A., Cieslak, M., Roalf, D. R., ... Bassett, D. S. (2021). Transdiagnostic dimensions of psychopathology explain individuals' unique deviations from normative neurodevelopment in brain structure [Publisher: Nature

- Publishing Group]. *Translational Psychiatry*, 11(1), 232. <https://doi.org/10.1038/s41398-021-01342-6>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85), 2825–2830. Retrieved August 19, 2025, from <http://jmlr.org/papers/v12/pedregosa11a.html>
- Pérez-García, F., Sparks, R., & Ourselin, S. (2021). TorchIO: A python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208, 106236. <https://doi.org/10.1016/j.cmpb.2021.106236>
- Price, B. H., Adams, R. D., & Coyle, J. T. (2000). Neurology and psychiatry [Publisher: Wolters Kluwer]. *Neurology*, 54(1), 8–8. <https://doi.org/10.1212/WNL.54.1.8>
- Romer, A. L., Ren, B., & Pizzagalli, D. A. (2023). Brain structure relations with psychopathology trajectories in the ABCD study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 62(8), 895–907. <https://doi.org/10.1016/j.jaac.2023.02.002>
- Royer, J., Kebets, V., Piguet, C., Chen, J., Ooi, L. Q. R., Kirschner, M., ... Bernhardt, B. C. (2024). Multimodal neural correlates of childhood psychopathology. *eLife*, 13, e87992. <https://doi.org/10.7554/eLife.87992>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sarrazin, S., Cachia, A., Hozer, F., McDonald, C., Emsell, L., Cannon, D. M., ... Houenou, J. (2018). Neurodevelopmental subtypes of bipolar disorder are related to cortical folding patterns: An international multicenter study [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/bdi.12664>]. *Bipolar Disorders*, 20(8), 721–732. <https://doi.org/10.1111/bdi.12664>
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., ... Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9), 3095–3114. <https://doi.org/10.1093/cercor/bhx179>
- SchizConnect*. (n.d.). Retrieved November 5, 2025, from <https://schizconnect.org/>
- Schulz, M.-A., Bzdok, D., Haufe, S., Haynes, J.-D., & Ritter, K. (2024). Performance reserves in brain-imaging-based phenotype prediction.

- Cell Reports*, 43(1), 113597. <https://doi.org/10.1016/j.celrep.2023.113597>
- Segal, A., Tiego, J., Parkes, L., Holmes, A. J., Marquand, A. F., & Fornito, A. (2025). Embracing variability in the search for biological mechanisms of psychiatric illness. *Trends in Cognitive Sciences*, 29(1), 85–99. <https://doi.org/10.1016/j.tics.2024.09.010>
- Shafiei, G., Esper, N. B., Hoffmann, M. S., Ai, L., Chen, A. A., Cluce, J., ... Satterthwaite, T. D. (2025). Reproducible brain charts: An open data resource for mapping brain development and its associations with mental health [Publisher: Elsevier]. *Neuron*, 113(22), 3758–3779.e6. <https://doi.org/10.1016/j.neuron.2025.08.026>
- Shalizi, C. R. (2025, February 8). *Advanced data analysis from an elementary point of view*. <https://www.stat.cmu.edu/~cshalizi/ADAFaEPoV/>
- Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., ... Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies [Number: 1 Publisher: Nature Publishing Group]. *Molecular Psychiatry*, 27(1), 281–295. <https://doi.org/10.1038/s41380-021-01161-7>
- Tamminga, C. (2014, December 9). Bipolar & schizophrenia consortium for parsing intermediate phenotypes (b-SNIP 1). <https://doi.org/10.15154/TNZS-A323>
- Tiego, J., Martin, E. A., DeYoung, C. G., Hagan, K., Cooper, S. E., Pasion, R., ... Fornito, A. (2023). Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. *Nature mental health*, 1(5), 304–315. <https://doi.org/10.1038/s44220-023-00057-5>
- Uhlhaas, P. J., Davey, C. G., Mehta, U. M., Shah, J., Torous, J., Allen, N. B., ... Wood, S. J. (2023). Towards a youth mental health paradigm: A perspective and roadmap [Number: 8 Publisher: Nature Publishing Group]. *Molecular Psychiatry*, 28(8), 3171–3181. <https://doi.org/10.1038/s41380-023-02202-z>
- UK biobank. (2025, January 10). Retrieved January 25, 2025, from <https://www.ukbiobank.ac.uk>
- Vanes, L. D., & Dolan, R. J. (2021). Transdiagnostic neuroimaging markers of psychiatric risk: A narrative review. *NeuroImage: Clinical*, 30, 102634. <https://doi.org/10.1016/j.nicl.2021.102634>
- Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, 180, 68–77. <https://doi.org/10.1016/j.neuroimage.2017.06.061>

- Varoquaux, G., & Colliot, O. (2023). Evaluating machine learning models and their diagnostic value. In O. Colliot (Ed.), *Machine learning for brain disorders*. Humana. Retrieved October 21, 2025, from <http://www.ncbi.nlm.nih.gov/books/NBK597473/>
- Varoquaux, G., Luccioni, S., & Whittaker, M. (2025). Hype, sustainability, and the price of the bigger-is-better paradigm in ai. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 61–75. <https://doi.org/10.1145/3715275.3732006>
- Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging [Publisher: Nature Publishing Group]. *Nature Neuroscience*, 20(3), 365–377. <https://doi.org/10.1038/nn.4478>
- Xu, B., Wang, H., Dall’Aglia, L., Luo, M., Zhang, Y., Muetzel, R., & Tiemeier, H. (2025). Beyond out-of-sample: Robust and generalizable multivariate neuroanatomical patterns of psychiatric problems in youth [Publisher: Nature Publishing Group]. *Molecular Psychiatry*, 30(6), 2525–2536. <https://doi.org/10.1038/s41380-024-02855-4>
- Zhao, Q., Nooner, K. B., Tapert, S. F., Adeli, E., Pohl, K. M., Kuceyeski, A., & Sabuncu, M. R. (2025). The transition from homogeneous to heterogeneous machine learning in neuropsychiatric research. *Biological Psychiatry Global Open Science*, 5(1), 100397. <https://doi.org/10.1016/j.bpsgos.2024.100397>

Appendix A

Methods details

A.1 Clinical datasets demographics

Table A.1 presents the demographics of the clinical datasets.

A.2 Child Behavior Checklist (CBCL) items for exploratory factor analysis

The CBCL comprises 118 problem items rated as "Not True (as far as you know)" (0), "Somewhat or Sometimes True" (1) or "Very True or Often True" (2) by parents. We excluded from analysis any item that was endorsed (rated 1 or 2) in less than 1% of the sample for any of the time points (baseline, +1 year, +2 years, +3 years or longitudinal). The 19 excluded items are 'Drinks alcohol without parents' approval' (cbcl_q02_p), 'Bowel movements outside toilet' (cbcl_q06_p), 'Cruel to animals' (cbcl_q15_p), 'Deliberately harms self or attempts suicide' (cbcl_q18_p), 'Hears sound or voices that aren't there' (cbcl_q40_p), 'Plays with own sex parts in public' (cbcl_q59_p), 'Plays with own sex parts too much' (cbcl_q60_p), 'Runs away from home' (cbcl_q67_p), 'Sees things that aren't there' (cbcl_q70_p), 'Sets fires' (cbcl_q72_p), 'Sexual problems' (cbcl_q73_p), 'Steals outside the home' (cbcl_q82_p), 'Thinks about sex too much' (cbcl_q96_p), 'Smokes, chews, or sniffs tobacco' (cbcl_q99_p), 'Truancy, skips school' (cbcl_q101_p), 'Uses drugs for non medical purposes (don't include alcohol or tobacco)' (cbcl_q105_p), 'Vandalism' (cbcl_q106_p), 'Wets self during the day' (cbcl_q107_p), 'Wishes to be of opposite sex' (cbcl_q110_p).

Then we combined together items with high polychoric correlation (> 75%)

Table A.1: **Demographics of clinical datasets** — ASD: Autism Spectrum Disorder, BD: Bipolar Disorder, SCZ: Schizophrenia. For the downstream tasks, "BD" and "Psychotic BD" labels were merged. Biological sex was encoded as binary in the datasets (M: male, F: female).

Study	Diagnosis	n	Biological sex (%F)	Age mean (std)	n sites
BIOBD	control	353	55.0%	40 (13)	8
BIOBD	BD	307	55.7%	40 (12)	8
BSNIP	control	199	58.3%	39 (12)	5
BSNIP	BD	117	65.8%	37 (12)	5
BSNIP	SCZ	192	30.7%	34 (12)	5
CANDI	control	29	41.4%	11 (3)	1
CANDI	BD	35	42.9%	10 (3)	1
CANDI	Psychotic BD	19	52.6%	12 (3)	1
CANDI	SCZ	20	45.0%	13 (3)	1
CNP	control	125	47.2%	32 (9)	1
CNP	BD	49	42.9%	35 (9)	1
CNP	SCZ	50	24.0%	36 (9)	1
PRAGUE	control	90	55.6%	28 (7)	1
PRAGUE	SCZ	43	44.2%	29 (6)	1
SCHIZCONNECT-VIP	control	329	47.1%	32 (13)	4
SCHIZCONNECT-VIP	SCZ	275	27.6%	35 (12)	4
ABIDE1	control	568	17.4%	17 (8)	20
ABIDE1	ASD	514	11.9%	17 (8)	20
ABIDE2	control	565	30.8%	15 (9)	17
ABIDE2	ASD	470	14.5%	15 (9)	19

resulting in the following 10 composites items:

- 'Distracted/Hyperactive': 'Can't concentrate, can't pay attention for long', 'Can't sit still, restless, or hyperactive', 'Inattentive or easily distracted'
- 'Attacks/Threatens': 'Cruelty, bullying, or meanness to others', 'Physically attacks people', 'Threatens people'
- 'Destroys': 'Destroys his/her own things', 'Destroys things belonging to

his/her family or others'

- 'Disobeys rules': 'Disobedient at home', 'Disobedient at school', 'Breaks rules at home, school or elsewhere'
- 'Peer problems': 'Doesn't get along with other kids', 'Gets teased a lot', 'Not liked by other kids'
- 'Weight problems': 'Overeating', 'Overweight'
- 'Stomach problems': 'Nausea, feels sick', 'Stomachaches', 'oddness': 'Strange behavior', 'Strange ideas'
- 'Rule breaking': 'Lying or cheating', 'Steals at home'
- 'Emotional dysregulation': 'Stubborn, sullen, or irritable', 'Sudden changes in mood or feelings',

A.3 Data augmentation parameters

Figure A.1 shows an example of random erasing and random cropping augmentations. The specific parameters used for data augmentation in the experiments are:

- **3D Random Erasing:** `RandomErasing` from `nid1` with parameters `prop=(0, x)` with `x` equals 0.2 (light data augmentation), 0.4 (medium) or 0.6 (strong), `ratio=(1, 5)`, `value='mean'`, `p=0.7`.
- **3D Random Crop** `RandomResizedCrop` from `nid1` with parameters `target_shape=(128, 128, 128)`, `scale=(0.08, 1)`, `ratio=(0.75, 1.33)`, `p=1`. These are the parameters used in SimCLR for 2D images (Chen et al., 2020).
- **Gaussian noise** `RandomNoise` from `TorchIO` with parameters `mean=0`, `std=0.1`.

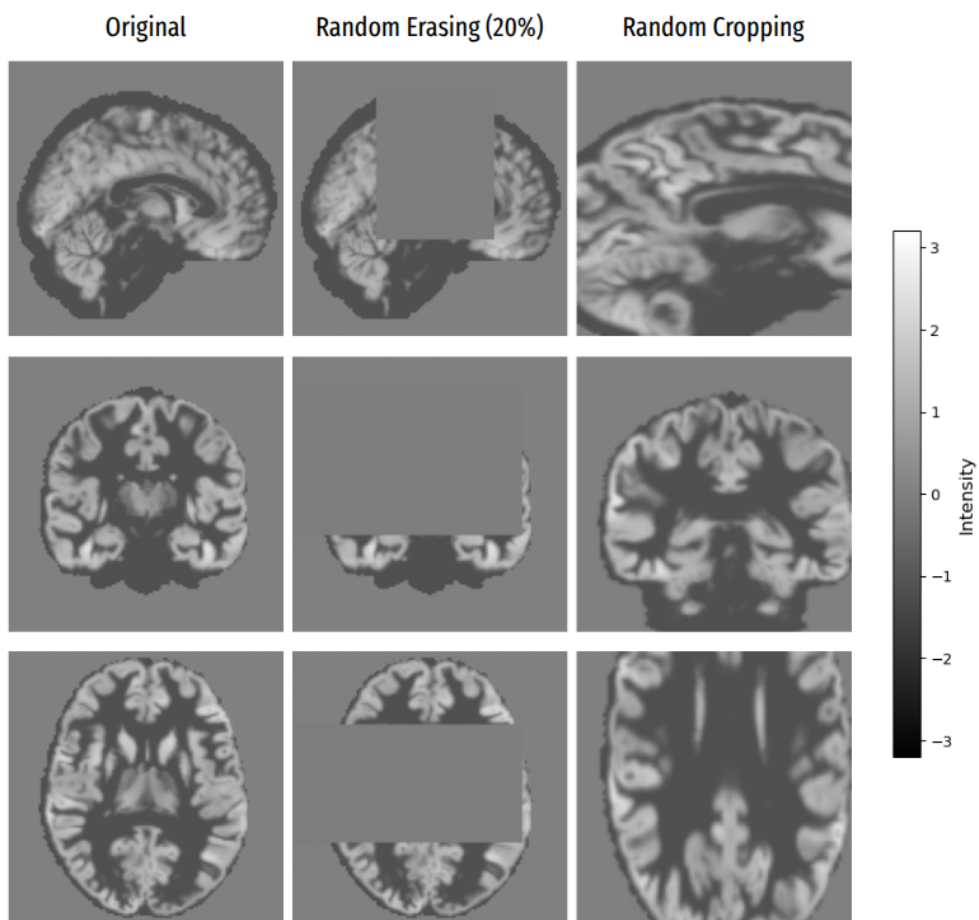


Figure A.1: **Data augmentations** — Sagittal, coronal and axial views of a preprocessed brain volume (left), with light random erasing (20% of volume erased) (center) or random cropping with default SimCLR parameters (right).

Appendix B

Additional results

Table B.1: **Loadings of the longitudinal 3-factor model** — Results of the EFA analysis on longitudinal CBCL data. Items are ordered by factor on which they primarily load and then by loading strength within this factor. The items that do not primarily load on one of the factors are shown at the end. The factors are internalising (Int.), externalising (F2) and neurodevelopmental (Ndev.).

	Int.	Ext.	Ndev.	primary factor
Unhappy, sad, or depressed	0.76	0.23	-0.13	F1
Too fearful or anxious	0.74	-0.09	0.12	F1
Worries	0.73	-0.06	0.05	F1
Feels too guilty	0.71	-0.05	0.05	F1
Feels worthless or inferior	0.70	0.24	-0.13	F1
Fears going to school	0.67	-0.02	0.01	F1
Feels he/she has to be perfect	0.66	0.00	-0.19	F1
Withdrawn, doesn't get involved with others	0.66	-0.01	0.14	F1
Self-conscious or easily embarrassed	0.64	0.05	0.04	F1
Too shy or timid	0.61	-0.19	0.10	F1
Fears he/she might think or do something bad	0.61	0.03	0.04	F1
Overtired without good reason	0.58	0.05	0.15	F1
Complains of loneliness	0.56	0.20	-0.00	F1
Nervous, highstrung, or tense	0.55	0.08	0.19	F1
Would rather be alone than with others	0.54	-0.00	0.11	F1

62 | Appendix B: Additional results

Sulks a lot	0.54	0.37	-0.07	F1
There is very little he/she enjoys	0.53	0.27	-0.04	F1
Feels dizzy or lightheaded	0.52	-0.08	0.13	F1
Underactive, slow moving, or lacks energy	0.51	-0.01	0.22	F1
Talks about killing self	0.50	0.34	-0.13	F1
Composite (stomach problems)	0.47	-0.07	0.16	F1
Refuses to talk	0.47	0.14	0.09	F1
Fears certain animals, situations, or places, other than school	0.47	-0.09	0.22	F1
Trouble sleeping	0.45	0.05	0.22	F1
Other (physical problems without known physical cause)	0.45	-0.07	0.17	F1
Secretive, keeps things to self	0.45	0.22	0.08	F1
Cries a lot	0.44	0.21	0.07	F1
Aches or pains (not stomach or headaches)	0.38	-0.04	0.21	F1
Headaches	0.37	-0.05	0.14	F1
Composite (attacks threatens)	0.05	0.82	-0.04	F2
Composite (disobeys rules)	0.01	0.78	0.09	F2
Gets in many fights	-0.02	0.77	0.05	F2
Argues a lot	0.07	0.77	0.00	F2
Doesn't seem to feel guilty after misbehaving	-0.02	0.71	0.10	F2
Temper tantrums or hot temper	0.15	0.70	0.00	F2
Teases a lot	-0.04	0.68	0.05	F2
Composite (destroys)	0.04	0.67	0.14	F2
Screams a lot	0.11	0.64	0.07	F2
Composite (rule breaking)	-0.03	0.63	0.18	F2
Swearing or obscene language	0.12	0.60	-0.01	F2
Easily jealous	0.23	0.59	-0.02	F2
Showing off or clowning	-0.16	0.57	0.28	F2
Demands a lot of attention	0.12	0.57	0.16	F2
Impulsive or acts without thinking	-0.06	0.56	0.40	F2
Bragging, boasting	-0.08	0.55	0.15	F2
Composite (emotional dysregulation)	0.38	0.53	0.03	F2
Hangs around with others who get in trouble	-0.08	0.52	0.20	F2
Composite (peer problems)	0.25	0.45	0.21	F2

Whining	0.18	0.37	0.18	F2
Poorly coordinated or clumsy	0.11	0.01	0.66	F3
Composite (distracted hyperactive)	-0.02	0.26	0.66	F3
Daydreams or gets lost in his/her thoughts	0.17	-0.08	0.61	F3
Stares blankly	0.19	-0.00	0.59	F3
Repeats certain acts over and over; compulsions	0.13	0.16	0.54	F3
Confused or seems to be in a fog	0.29	0.04	0.52	F3
Composite (oddness)	0.17	0.21	0.52	F3
Talks too much	-0.09	0.29	0.46	F3
Nervous movements or twitching	0.21	0.03	0.46	F3
Acts too young for his/her age	-0.01	0.26	0.45	F3
Speech problem	-0.03	-0.01	0.44	F3
Fails to finish things he/she starts	0.11	0.30	0.44	F3
Gets hurt a lot, accident prone	0.14	0.04	0.44	F3
Prefers being with younger kids	0.10	0.15	0.40	F3
Picks nose, skin, or other parts of body	0.03	0.13	0.39	F3
Stores up too many things he/she doesn't need	0.17	0.15	0.36	F3
Feels or complains that no one loves him/her	0.49	0.52	-0.19	
Feels others are out to get him/her	0.42	0.47	-0.06	
Unusually loud	-0.03	0.46	0.39	
Suspicious	0.35	0.44	0.11	
Poor school work	0.05	0.32	0.38	
Can't get his/her mind off certain thoughts; obsessions	0.32	0.17	0.37	
Clings to adults or too dependent	0.35	0.07	0.31	
Sleeps more than most kids during day and/or night	0.34	0.08	0.12	
Doesn't eat well	0.33	0.13	0.13	
Talks or walks in sleep	0.04	-0.01	0.31	
Constipated, doesn't move bowels	0.31	-0.00	0.19	
Nightmares	0.30	0.07	0.30	
Sleeps less than most kids	0.29	0.12	0.24	
Vomiting, throwing up	0.28	-0.06	0.24	
Problems with eyes (not if corrected by glasses)	0.18	-0.01	0.26	

64 | Appendix B: Additional results

Rashes or other skin problems	0.26	-0.06	0.19
Wets the bed	-0.04	0.10	0.25
Prefers being with older kids	0.13	0.25	0.20
Bites fingernails	0.09	0.07	0.24
Composite (weight problems)	0.21	0.12	0.14
Thumb-sucking	0.01	0.04	0.20

TRITA-EECS-EX-2026:85
Stockholm, Sweden 2026

www.kth.se