# Machine learning for neurodevelopmental disorders

Clara Moreau, Christine Deruelle, Guillaume Auzias

Book: Machine Learning for Brain Disorders

# Chapter 31

# “Machine learning for neurodevelopmental disorders”

Clara Moreau[1], Christine Deruelle[2], Guillaume Auzias[*,2]

[1] Human Genetics and Cognitive Functions, CNRS UMR 3571, Université de Paris, Institut Pasteur, 25 rue du Dr. Roux, Paris, France

[2] Aix-Marseille Université, CNRS, Institut de Neurosciences de la Timone, UMR 7289, Marseille, France.

**Running head:** Neurodevelopmental disorders

***Corresponding author:**
Guillaume Auzias
Institut de Neurosciences de la Timone
Faculté de Médecine
27, boulevard Jean Moulin
13005 Marseille - France
E-mail: guillaume.auzias@univ-amu.fr

# Abstract length: maximum of 250 words.

Neurodevelopmental disorders (NDDs) constitute a major health issue with >10% of the general worldwide population affected by at least one of these conditions - such as Autism Spectrum Disorders (ASD) and Attention Deficit Hyperactivity Disorders (ADHD). Each NDD is particularly complex to dissect for several reasons, including a high prevalence of comorbidities and a substantial heterogeneity of the clinical presentation. At the genetic level, several thousands of genes have been identified (polygenicity), while a part of them was already involved in other psychiatric conditions (pleiotropy). Given these multiple sources of variance, gathering sufficient data for the proper application and evaluation of machine learning (ML) techniques is essential but challenging. In this chapter, we offer an overview of the ML methods most widely used to tackle NDDs complexity - from stratification techniques to diagnosis prediction. We point out challenges specific to NDDs such as early diagnosis, that can benefit from the recent advances in the ML field. These techniques also have the potential to delineate homogeneous subgroups of patients that would enable a refined understanding of underlying physiopathology. We finally survey a selection of recent papers that we consider as particularly representative of the opportunities offered by contemporary ML techniques applied to large open datasets, or that illustrate the challenges faced by current approaches to be addressed in the near future.

# Keywords

Neurodevelopmental disorders, Autism Spectrum Disorders, Attention Deficit Hyperactivity Disorders, Machine learning, Pattern recognition, Classification, Clustering, Stratification

# 1. A brief introduction to neurodevelopmental disorders

Neurodevelopmental disorders (NDDs) cover a large range of pathologies. This term can be used to refer to known genetic syndromes such as Fragile X syndrome or, in a much broader sense, include conditions with multifactorial etiology such as Autism Spectrum Disorders (ASD), Attention Deficit Hyperactivity Disorders (ADHD), or developmental dyslexia. Even more broader are the definitions from the DSM-5 or the ICD10 which also encompasses intellectual disabilities (ID), communication disorders, specific learning disorders, motor disorders [1]. NDDs embrace defects that disturb the developmental function of the brain, which could lead to neuropsychiatric complications, learning difficulties, language or non-verbal communication problems, or motor function disabilities. However, although there is a tight intrication between NDDs and psychiatric disorders - for whom manifestations come later in life - phenomenological categories used in the adult population do not apply consistently in NDDs. The latters are conditions for which the cause or the onset is located during gestation or birth and should be distinguished from late-onset disorders. We refer to [2]–[4] for an historical view of the standardized tools allowing for reliable and valid categorical distinctions, available to the community since the 2000s.

NDDs constitute a critical health problem in our society. More than 10% of the general worldwide population is affected by neurodevelopmental disorders [5]. The consequences of NDDs impact a person's lifetime, so patient management represents a major cost for society. Important healthcare advances have improved the life course of several NDDs (e.g. very low birth weight preterm infants, congenital hydrocephalus) and extended the expected lifespan of others (e.g. cystic fibrosis). The assessment and study of individuals with NDDs become thus an increasingly crucial issue. Researchers and clinicians have strongly emphasized the importance of early identification and intervention to improve the level of functioning. However, because of the high complexity intrinsic to these pathologies, we face a lot of misdiagnoses or even missed diagnoses which prevent early and effective therapeutic interventions. As an illustration, ⅕ of children diagnosed with ADHD or ASD in the population are currently misdiagnosed, which leads to a failure to get the adequate treatment or the administration of an unnecessary one.

NDDs are particularly complex to approach and to diagnose for several reasons. First, comorbidities are common in NDDs. Comorbid clinical features have been shown to be the rule rather than the exception in NDDs, adding to the complexity of proper diagnostic boundaries delineation. Over a third of individuals with ASD meet criteria for ADHD, Obsessive Compulsive Disorder (OCD), disruptive behavior disorders, anxiety and mood disorders, intellectual disability, or epilepsy, inducing various diagnostic combinations [2], [6], [7]. This overlap across conditions probably originates from a shared neurological etiology. As a consequence, studies that exclude other psychiatric disorders have limited translational application because of the pathophysiological overlap between many comorbid disorders (see figure 1 for an illustration of this issue).

In relation to this first issue, neurodevelopmental disorders overlap a lot in terms of etiology because of important epidemiological comorbidity and community of symptoms [8]. NDDs show indeed considerable overlap both neuropsychologically, physiologically, and genetically. For instance, the presence of certain behavioral characteristics, such as attention problems do not systematically indicate a specific diagnostic entity (e.g. ADHD) but instead, attention problems occur across a large variety of disorders (such as in ASD or in anxiety disorders). When biological bases are considered, the level of heterogeneity remains elevated. A wide range of neurological substrates have been associated with individual disorders. For example, ADHD has been associated with differences in gray matter within the anterior cingulate cortex, caudate nucleus, pallidum, striatum, cerebellum, prefrontal cortex, the premotor cortex, and most parts of the parietal lobe [9].

Similarly, at the genetic level, both common as well as rare, and structural as well as sequence, variations have been identified as contributing to NDDs. There are multiple examples in which the identical variant has been found to contribute to a wide range of formerly distinct diagnoses, including autism, schizophrenia, epilepsy, intellectual disability and language disorders. These include variations in chromosomal structure at 16p11.2, rare *de novo* point mutations at the gene *SCN2A*, and common single nucleotide polymorphisms (SNPs) mapping near loci encoding the genes *ITIH3, AS3MT, CACNA1C,* and *CACNB2*. In the case of autism, high genetic heritability (70-80%) with more than 1000 genes contributing to ASD has been yielded [10]. These selected examples point that heterogeneity in these pathologies is clearly multidimensional [3]. As a result, conferral of a diagnosis based on DSM-5 or ICD-10 criterion ascribes an underlying cause to the various behavioral difficulties

without a method available to verify that the disorder arises from underlying biological dysfunction.

The specificity of NDDs relative to psychiatric disorders (covered in **Chapter 32**) is that the challenges induced by the intrication of a spectrum of conditions are potentialized by the developmental dimension. Indeed, the developmental transformation is a major contributor to the multidimensional heterogeneity across individuals affected by NDDs. Brain developmental trajectory exhibits marked variations across individuals [11], [12], but also across brain regions [13], [14]. The development course concerns cognitive, neuronal, epigenetic maturation processes that follow distinct, yet inter-dependent nonlinear trajectories [15], [16]. During development, reorganization and competition for function are highly active. Compensatory mechanisms can thus interfere with potential alterations of the nervous system in individuals with NDDS. The timing of these alterations is of high relevance as different neural systems are selectively vulnerable to injury at different phases of prenatal and post-natal development [17]. This plasticity partially explains the heterogeneity in behavioral and cognitive dysfunction associated with early alteration, ranging from subtle to diffuse and profound. In addition, the functional impairments can be observed immediately in some individuals while in others the full range of deficits may not manifest until later in life [18].

As a consequence, early diagnosis is key since early medical intervention would benefit from the remarkable plasticity of the immature brain, allowing the patient to adapt and/or develop compensatory mechanisms. On the basic research side, investigating earlier allows to reduce the influence of compensatory mechanisms and secondary perturbations. Studies focused on young children are more likely to reach the causes, whereas in adult populations consequential or adaptation abnormalities likely contaminate the observations.

There are thus crucial needs in NDDs for a better detection of early, subtle signs of neurodevelopmental pathology, and more accurate prediction of the evolution of the impairments. Gaining insight on the pathophysiological processes and the identification of more homogeneous subtypes is also required for the identification of new targets for drug development.

To address these needs, collective efforts have been made to constitute large public datasets giving access to sufficient amounts of multi-dimensional data covering the dimensions mentioned above (see e.g. [19]). Recently, we have witnessed the constitution of

large databases trying to address these issues and which we will refer to in the following chapters. We can mention for instance, ABCD [20], ABIDE [21], EU-AIMS [22], ADHD200 [23], (see **Chapter 24)** for general considerations regarding the rise of openly accessible large datasets. It induced a crucial need for statistical approaches tailored for the data-rich setting and thus called for closer collaboration with the field of Machine Learning.

Unsurprisingly, the NDDs having the largest prevalence, and thus, the greater societal impact and the easier recruitment, are largely overrepresented in these databases. As a consequence, they are also overrepresented in the literature of ML techniques applied to NDDs. In the remainder of this chapter, we focus on ASD and ADHD. With regard to the characteristics mentioned above, we argue that ASD and ADHD are highly representative of the NDDs in general. As detailed in **boxes 1 & 2**, they are the two most common neurodevelopmental disorders observed in childhood, and they present considerable variability, both within and across conditions. These two syndromes share most of their comorbidities, while 40 to 83% of children with ASD also have ADHD [24], 28 to 87% of children with ASD showing symptoms of ADHD [25]. See [26] for a comparison of the outcomes from recent neuroimaging studies in these two disorders. As a consequence of this heterogeneous clinical presentation, we clearly face a lack of objective criteria for diagnosis for these two disorders as well as for the other NDDs.

**Box 1: Autism spectrum disorder (ASD).**

ASD is a complex neurodevelopmental condition with lifelong impacts. Current prevalence is estimated to be at least 1.5% in developed countries. The male-to-female ratio is estimated to 4:1 in this pathology. This sex ratio varies, however, according to intellectual disability (ID). Reported median sex ratios of 6:1 among normal-functioning subjects and 1.7:1 among cases with moderate to severe ID [27]. Individuals with ASD suffer from a specific combination of deficits in social communication and repetitive behaviors, severely restricted interests and sensory behaviors from early in life. Despite the vast resources devoted to the study of ASD, its pathogenesis remains largely unknown. Recent genetic studies have identified a number of rare de novo mutations and provided insight into polygenic risk, epigenetics, and gene-by-environment interaction related to autism or autistic traits [28]. In addition, epidemiologic investigations focusing on nongenetic factors have identified advanced parental age and preterm birth as risk factors for ASD and have suggested that prenatal exposure to air pollution and short inter pregnancy interval are also potential risk factors. See e.g. [29] for more detailed information.

**Box 2: Attention Deficit Hyperactivity Disorder (ADHD).**

ADHD is one of the most common neurodevelopmental disorders, characterized by inappropriate and developmentally harmful levels of inattention, hyperactivity, and impulsivity. It affects boys more often than girls. Its prevalence in the general population is between 3 and 4%. ADHD is diagnosed according to strictly defined criteria but there is still no fiable biomarker of the pathology. The causes of ADHD are complex and multifactorial, with genetics, early environment and gene-environment interplay being involved. Although ADHD is highly heritable, and multiple types of genetic variants are associated with the disease, none of them can be used as diagnostic. Diagnostic thresholds are given by both the ICD-10 and the DSM-5, but the clinical features of ADHD behave as continuously distributed dimensions and vary considerably between individuals. Clinical features are heterogeneous. ADHD profiles include not only its definite symptoms (hyperactivity-impulsiveness, inattention) and features of other neurodevelopmental disorders but also additional cognitive deficits such as impaired working memory and planning. Early comorbidity with developmental, learning, and psychiatric problems such as ASD, is very frequent. ADHD is lifelong but its course and outcome are highly variable. Core symptoms such as the hyperactivity observed at preschool age may turn into inattention and executive dysfunction in older children for instance. See e.g. [30] for further information.

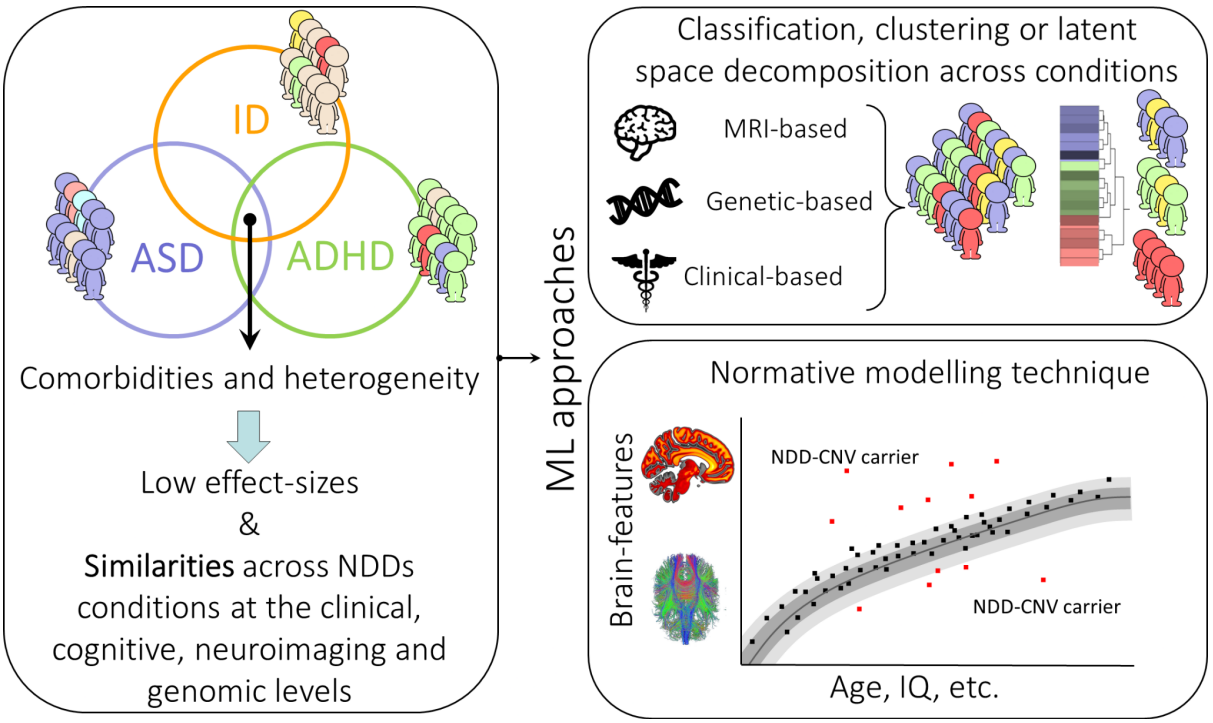


Figure 1 Left: As introduced in section 1, the complexity of NDDs comes from the combination of multiple sources of heterogeneity acting at different levels, and that overlap across conditions as illustrated here with ASD, ADHD, and intellectual disability (ID). Right: As described in section 2, ML approaches are instrumental to characterize and overcome the heterogeneity at each level with dedicated techniques.

# 2. What are the main challenges in these conditions that can be addressed using machine learning?

Given these multiple sources of variance, gathering sufficient amounts of data for proper application and evaluation of machine learning (ML) techniques is essential, but also very challenging. As underlined earlier and illustrated on Figure 1, NDDs, and more specifically the two we focus on, present a number of specific challenges that can be formulated in terms of heterogeneity, trajectory of development, and comorbidities.

In this section, we give an overview of the methods most widely used in the NDDs literature and point to specific challenges that can benefit from the recent advances from the ML field. We refer readers interested in an exhaustive view of the available approaches and their performances in the context of NDDs to the following recent review papers [31]–[36]. We organize this overview by following the historical evolution of the methods used in the field. The first applications of ML techniques were focused on classification tasks. Indeed, classification techniques can be designed for the prediction of later evolution and are thus in principle well suited to address the challenge of early diagnosis. We then observe a progressive shift towards regression, latent space decomposition, and stratification purposes. These approaches have the potential to uncover more homogeneous subpopulations of patients that would enable refined understanding of underlying physiopathology. More recently, specific approaches have been proposed for characterizing the atypical brain maturation trajectory in NDDs. Finally, we discuss the potential of deep learning techniques for learning representations that might represent a major step towards prediction at the individual level, which is crucial for translation into clinical applications.

## 2.1. The classical analysis approach failed to reach consensus

Historically, the classical analysis approach consisted in designing a study starting from the definition of an ‘atypical’ population of interest, based on particular clinical scores selected among the behavioral assessments used for diagnosis. This population of interest is compared to a group of control subjects, following a feature defined a priori such as ‘the volume of a specific cortical region estimated from anatomical MRI’. As extensively described in e.g. [37]–[39], this corresponds to statistically testing the hypothesis: Does the atypical population differ, on average, from controls in the selected feature? Statistically speaking, this amounts to a case-control study using univariate hypothesis testing for one or a few features. The large literature of early studies following this approach allowed to refine the characterization of the different sources of heterogeneity presented above, and shed light on the lack of biological validity of categorical representations of NDDs that manifest in the evolution of the nosology, for instance moving from ‘autism’ to ‘autism spectrum disorders’ [3]. However, as we progressed in our understanding of the interactions between genetics, biological brain and behavior, the limits of the group statistics and univariate approaches became blattent:

- *Limitations of classical univariate analysis techniques.*

  The univariate approach is prevalent in the literature for historical reasons. It relies on the implicit assumption that different brain regions and/or different features are independent, while more and more evidence supports the opposite view: effects are spread across several brain regions, possibly located far from each other. Knowing the various sources of variance in NDDs data described earlier, it is unlikely that a single feature may capture a large portion of that variation and thus be interpreted in terms of underlying biological processes. It is thus not surprising that the effect sizes reported in meta-analyses remain small. In addition to potentially reduced statistical power, the problem of inflated false discovery rate in univariate analysis framework has been raised and extensively discussed [40]. Multivariate approaches are much more relevant in this context. Indeed, combining in a multivariate approach a group of features having small effect size when considered independently might lead to a large effect [38].

- *Limitations of group statistics.*

  As extensively discussed in [41], group statistics all focus on first order statistics (group means), thereby seeking a pattern of atypicality that is consistent across the population (i.e. the 'average patient'). Indeed, mean group differences may reflect a systematic shift in the distribution of the clinical group and thus provide useful information on altered processes in that population. However, those differences do not delineate variability within groups [38]. In addition, the evolution of the DSM by regrouping conditions that were considered in previous versions as distinct (e.g. Asperger and pervasive developmental disorders not otherwise specified) induced an increase in the heterogeneity of the populations included in studies on ASD [37]. Group comparisons based on diagnosis thus present the major caveat of ignoring psychiatric comorbidities, which are common in NDDs. It thus becomes obvious that group statistics applied to populations defined based on diagnostic categories are inadequate. Indeed, categorical diagnoses from the DSM are increasingly found to be incongruent with emerging neuroscientific evidence that points towards shared neurobiological dysfunction underlying NDDs [42]. See e.g. [39] for extensive discussions on the limitations of the diagnostic-first approach in comparison to the alternative strategy that begins at the level of molecular factors enabling the study of mechanisms related to biological risk, irrespective of diagnoses or clinical manifestations.

The combination of univariate statistics and mean group difference analysis applied to heterogeneous populations with small sample sizes resulted in highly inconsistent findings. Indeed, most of the published findings are not consistent and were not replicated. The recent challenge [43] further illustrates the intrinsic limitation of the group statistics framework, but also that state of the art ML techniques do not systematically outperform classical approaches in such a binary classification task. In this context, deep-learning techniques were prone to overfitting with poor generalization to unseen dataset, while simpler approaches had a stable prediction performance when applied to new data. It is important to stress that several limitations from this early literature do fully apply to more advanced ML techniques and/or multivariate data analysis strategies. While the problem of inflated false discovery rate in univariate analysis framework has been extensively discussed [40], the problems related to improper evaluation and validation of ML techniques (e.g. overfitting and biases induced by inadapted cross validation strategy or absence of a truly independent test set) emerges in the recent literature [31], [44]–[46]. While discussing the limitations of cross validation for estimating the potential overfitting of statistical models is beyond the topic of this chapter, we stress the crucial importance of raising awareness of these aspects. We refer interested readers to essential guidelines and recommendations that have been provided in [43], [47]–[51]. Indeed, uncovering potential biases in the models validation strategy is a tedious but essential step. [52] is a nice illustration of the major gains in interpretation resulting from an extensive analysis of the most influential factors.

## 2.2. Promises of ML in NDDs

The rise of big data and the sustained advances in ML enable in principle the integration of various and heterogenous characteristics such as behavioral profiles, imaging phenotypes, and genomics. The extraction and manual construction of features from each data type, also termed as **feature engineering**, did undergo continuous progress in tight relation with innovations in the acquisition processes. As an illustration, the imaging phenotype today covers a wide range of features extracted mainly from MRI data. For instance a variety of measures can be extracted from diffusion weighted imaging [53], from basic estimation in each voxel such as the *fractional anisotropy* to higher level connectivity measures in each anatomically-defined *fiber tract*, or even connections between distant anatomical regions (structural connectivity). On the genetics side, polygenic risk scores (PRS) are additive models developed to estimate the aggregate effects of thousands of common variants with very small individual effects. They can be computed for any individual to estimate the

risk/probability for a particular trait conferred by common variants [54]. Feature engineering is a crucial step in the analysis since the biological relevance of the features directly impacts the interpretation, and the strategy used to manage potential interaction across different features might determine the performance of the analysis procedure more than the ML algorithm itself. In parallel, the increase in the size of the available data enables the training of more complex algorithms, making it possible to investigate central questions related to the dynamics of normal and abnormal development by means of advanced ML techniques.

## 2.3. Classification and prediction: supervised learning for NDDs

Classification techniques consist in learning a model allowing to separate different groups of subjects based on a set of training data that have been labeled and are thus subtypes of *supervised* machine learning techniques. In this context, classification techniques integrate biological and/or behavioral measures in order to extract a *predictive pattern* corresponding to the diagnosis. Classification techniques used in the literature of NDDs are the same as those used in the field of psychiatry and span the whole range of methods detailed in **Chapters 1 to 6**, from simple linear models to most recent deep networks. References [31], [32], [34], [51], [55] provide a detailed overview of the recent applications of classification techniques in the context of ASD and ADHD. The general trends indicate that Linear Discriminant and Logistic Regression Classifiers were prominent until around 2014, most studies focusing on a single modality (usually structural or functional MRI). Support Vector Machines (SVM) then became the most commonly used approach due to their performance in the small sample-high dimension regime but also their ability to perform non-linear classification. Approaches based on ensembles of classifiers were more recently developed to combine data from several modalities or acquired in different settings (e.g. different scanners). Even more recently, deep learning techniques neural networks were applied to populations of a few hundred subjects. We will discuss the potential of these advanced approaches later, in a dedicated section. In terms of input data types, structural and functional MRI modalities are overrepresented in comparison to diffusion MRI, EEG, and behavioral data. Classification techniques based on genetics are getting more and more attention (e.g. using polygenic risk scores). Due to the complex and specific data preprocessing required for each modality (see e.g. [43]), combining features extracted from several modalities into a multimodal classification technique represents important additional

challenges. Only a few studies did explore the potential of combining several modalities so far (e.g. 4 studies among 57 reviewed in [31]), but the initiatives for sharing preprocessed data such as those in [23], [56] will facilitate this type of analyses in the future. Multimodal classification techniques did not demonstrate major performance gain so far, but further improvements can be expected by better exploiting the complementarity of the information across different modalities [32]. In terms of classification performances, the high accuracy (>80%) reported in early studies tended to decrease while sample size increased [31], [32], suggesting that the impressive results obtained on small cohorts were affected by overfitting, sampling biases and artificially reduced heterogeneity within and across the populations involved. Note that the decreasing effect sizes of group comparison studies might also be related to the evolution in the definition of autism toward a more inclusive and heterogeneous population [57].

In parallel with this decrease with time in the performance, the research field on psychopathology did initiate a shift, moving away from diagnostic categories based on symptoms to the concept of dimensions related to more objective measures and having better cognitive and biological validity. In particular, the US National Institute of Mental Health initiated in 2009 the Research Domain Criteria (RDoC) project to develop a classification system for mental disorders based upon fundamental dimensions of neurobiology and observable behavior that cut across current heterogeneous disorder categories [58], [59]. Of note, this research classification system diverges from one intended for routine clinical use in multiple respects [60]. Following this progressive conceptual shift, the major methodological challenge to be addressed moved away from classification and diagnostic prediction to latent space decomposition and stratification.

## 2.4. Latent space decomposition and clustering: unsupervised learning for NDDs

Following the progressive confirmation of the inadequacy of mutually-exclusive diagnostic categories, behavioral assessment for quantifying ASD traits in any given individual were introduced, such as the autism spectrum quotient questionnaire [61] and the Social Responsiveness Scale (SRS) [62]. A number of studies used these scores to demonstrate that ASD traits are also present in the typically developing population as well as in other

NDDs such as ADHD [63]. These studies supported the view of a continuum across NDDs, and emphasized the need for novel approaches to identify general psychopathology dimensions that cut through diagnostic boundaries. Such data-driven dimensions would ultimately enable the identification of new targets for treatment development and to stratify the NDDs in subgroups more appropriate for treatment selection [58], [59], [64]. Uncovering the hidden intrinsic structure in the data is a well known ML problem that has been formulated as *unsupervised* learning in opposition to *supervised* learning tasks such as classification where the algorithm learns to predict a label based on a training set for which the true label is known (see **Chapters 1-2** or e.g. [65]). Unsupervised ML techniques consist in fitting a statistical model to the data by implementing specific assumptions regarding the relationships between the input features and on the supposed hidden structure. A general assumption to all unsupervised techniques is that there exists a non-negligeable degree of correlation across some of the features in the actual data, which justifies the search for a more compact optimal representation. Depending on the assumptions regarding the hidden structure to discover, unsupervised techniques can be divided in two classes: latent space decomposition and clustering. Latent space decomposition techniques aim at projecting the data onto a new feature space of lower dimension in which a large portion of the variance can be explained by a few factors. The underlying assumption is that the projected features vary continuously along the axes of this compact subspace. In contrast, clustering techniques seek to partition the data into distinct groups (often termed as population stratification) so that the observations within each group are similar to each other, while observations in different groups differ from each other. The underlying assumption is thus that a categorical representation is more appropriate than in the case of the latent space decomposition approach. In contrast with the classification task, the algorithm is designed in this case to identify homogeneous subpopulations within and across diagnostic categories. Several recent approaches propose a unified framework combining the advantages of both the dimensional and categorical models [3], [66], [67].

All unsupervised approaches face two main challenges in the context of NDDs. First, since we are dealing with a limited amount of data, the number of dimensions or clusters that can be identified needs to remain limited in order to avoid the curse of dimensionality i.e. when an infinite number of solutions can fit data equally well [64], [65], [68]. As a consequence, in the majority of studies, the set of input features (and thus the dimension of the input space) is selected based on data availability or prior knowledge, which raises the problem of establishing an optimal set of variables of particular relevance for NDDs [69]. Automated

feature selection procedures can be used to reduce the dimensions to be explored (see [69] for a recap of the approaches explored so far in ASD), but the fundamental problem of limited amount of data relative to the very large dimension to explore remains [64]**.** The second major challenge is the validation, since with unsupervised approaches no ground truth data is available by definition, unlike in the case of supervised ML. The relevance of the resulting dimensions or clusters should be assessed in terms of interpretability relative to external measures, that would ideally have some clinical relevance. Replication on a fully independent dataset allows to assess the generalizability and reduces the risk of overfitting. This is however very hard to achieve since the number of datasets available with identical measures is limited. As a consequence, it is crucial to keep in mind that unsupervised learning is only meaningful in relation to some context [70]. As extensively discussed in [64], “due to the vast dimensionality of the human population (based on environment, behavior, biology/physiology, etc.) there are multiple ways that the population might be subcategorized that are valid and ‘real’; however, any given subgrouping might not be important for the question we care about.”

Contrary to the classification task where the literature is very rich, latent space decomposition and stratification studies in NDDs are emerging approaches and only few findings have been published so far. Two recent publications review unsupervised approaches applied to neuroimaging in the context of ASD: [31] covered 19 studies published since 2018, and [69] identified 12 studies among which two were already included in [31]. For an extensive review covering the literature back to 2001, see [71]. The methods used range from the most common such as principal component analysis for latent space decomposition and K-means for clustering, to more advanced techniques such as nonnegative matrix factorization, spectral clustering, Gaussian mixture models, and Bayesian latent factor analysis such as Indian Buffet Processes. Most advanced approaches such as Bayesian latent factor analysis techniques enable to infer the number of latent factors and number of putative subpopulations from the data, and can be interpreted both in terms of categorical and dimensional aspects of the heterogeneity in NDDs [69], [72]. On the genomics side, multivariate approaches such as canonical correlation analysis, partial least square regression are the tools of choice for investigating the relationship between genomic variants, neuroimaging features, psychiatric conditions, and behavioral traits [39]. The development of specific methods allowing to better model the multivariate genetic covariance structure in genome-wide association studies is a very active field. For instance, [73] introduced a new approach called genomic structural equation modeling, that allows to

investigate shared genetic effects across phenotypes, while concurrently testing for causes of divergence. Importantly, this evolution in the methods reflects the progressive integration of latent space decomposition and clustering techniques into unified approaches. A promising avenue of research that benefited from access to larger datasets in the past years consists in combining neuroimaging and genomics. Indeed, the effects of latent factors derived from genomics on neuroimaging endophenotypes demonstrate higher reproducibility and larger effect-size than in the previous literature [39], [74].

In terms of evaluation and performances, the studies are highly dependent on the data and the assumptions that are made, either implicitly or explicitly. An illustration of this dependency on the application is the variation in the number of subtypes reported, ranging from two to six across the neuroimaging studies on ASD included in the two reviews [31], [69]. In [71], the authors cover a much broader literature (159 articles) by relaxing inclusion criteria compared to the two others. This exhaustive review identifies seven validation strategies, defined as follows: “cross-method replication”, “subtype separation”, “independent replication”, “temporal stability”, “external validation”, “parallel validation”, and “predictive validation”. They provide the distribution of the number of identified subtypes across the reviewed studies, with a range of values varying between 1 and 16, but 82% of all studies reporting between two and four subtypes. Of note, this review underlies as major challenges the access to large and multidimensional datasets, and the design of an unbiased validation framework. We refer interested readers to [71], in particular for the didactic description of the various validation strategies that apply to the literature of ASD and more generally to psychiatry or other clinical groups.

## 2.5. Normative modeling for NDDs

Normative modeling gained great interest in the context of psychiatry recently, and the first applications to NDDs confirm the particular relevance of this approach in this context. [75] introduced normative modeling as an alternative to clustering for parsing heterogeneity across the full range of population variation, i.e. spanning both clinical and healthy cohorts. In the approach proposed by [75], the normative models were estimated using Gaussian process regression [76]. The flexibility of this Bayesian method enables to define a mapping between any quantitative biological measures and clinically relevant variables, and offers desirable properties such as robustness to over-fitting, and principled ways for tuning hyper-parameters. Gaussian process regression is flexible but does not scale with an

increase in sample size. More importantly this technique can lead to inaccurate uncertainty estimates when the data are non-Gaussian [77]. Less demanding alternative approaches have been proposed. In [78], the authors used a non-parametric local weighted regression to fit a smooth curve through data points. Based on the assumption that the estimated regression is likely to be smooth, [79] proposed to estimate nonlinear effects using a smoothing spline model. This approach is a special case of Gaussian process regression. It is thus less adaptive, but presents a lower computational cost than Gaussian process regression. [80] presented a novel framework based on spline interpolation combined with likelihood warping and Bayesian estimation that allows to scale normative modeling to big data cohorts. Another approach based on generalized additive models was proposed in [81], [82]. The very last version of normative models was presented recently by [83] with the generalized additive models for location, scale and shape (GAMLSS), a flexible modeling framework that can model heteroskedasticity, non-linear effects of variables, and hierarchical structure of the data. As demonstrated in [84] with features extracted from more than 120,000 MRI, these models can be estimated on very large datasets. They are however not suitable for small datasets since the higher flexibility of such a model would be detrimental and might lead to overfitting.

Normative models are highly relevant for analyzing neuroimaging data since they can be fit at each brain location to estimate regional specificity. In the context of NDDs, two advantages are particularly critical. First, normative modeling is efficient to disentangle the effects related to brain maturation dynamics and neurodevelopmental diseases in a data-driven way. Indeed, the Bayesian framework enables estimating distinct variance components. The effect of age within the reference cohort is estimated by non-linear interpolation, which is appropriate in this period of highly active neurodevelopment (Batalle, Edwards, and O'Muircheartaigh 2018; Thompson et al. 2020).

Second, normative modeling provides uncertainty measures to quantify the variation across the estimated mean within the reference cohort and the deviation of each patient from the group mean. This enables the detection and mapping of subject-specific patterns of abnormality in each individual. The statistical inference at the level of the individual participant is the key to explicitly characterize the heterogeneity underlying clinical conditions. It represents a concrete alternative to the limitations of the case-control analysis seeking a pattern of atypicality that is consistent across the population as discussed in section 2.1. In the normative modeling framework, a deviation map is computed for each

individual based on extreme values statistics, which does not require that atypicalities overlap across participants. These individual deviation maps can then be analyzed (e.g. using unsupervised ML approaches described in section 2.4) to identify distinct patterns of abnormality, i.e. to characterize putative subpopulations.

See [41], [83], [86], [87] for further description of the normative modeling framework, and recommendations to guide future applications. The release of two python packages contributed to the widespread use of this approach: https://github.com/ppsp-team/PyNM and https://github.com/amarquand/PCNtoolkit. A didactic tutorial with a step-by-step comparison of the different normative modeling approaches on synthetic data illustrating their advantages and limitations is available online here: https://github.com/ppsp-team/PyNM/tree/master/tutorials.

## 2.6. Potential and challenges of deep learning

Deep learning is a class of ML algorithms characterized by their specific internal architecture as multi-layered neural networks. These multiple layers enable the striking capacity to progressively extract higher-level features without extensive priors injection. Their advantages compared to previous approaches are of crucial importance in a large range of applications and explains the considerable attention gained by DL in the wider scientific community. See e.g. [88] for a detailed description of the DL methods used in the literature to investigate the neuroimaging correlates of psychiatric and neurological disorders. Conceptually, DL techniques are particularly relevant for the investigation of NDDs for the following reasons:

- *Integrated learning of hierarchy of features.* As mentioned in section 2.2, classical ML algorithms leverage sets of structured features extracted from the input data. This feature engineering step relies on a priori regarding the data and has a strong influence on the performances. DL algorithms process directly the raw data without requiring prior feature extraction. During the learning, the algorithm can determine the optimal hierarchy of most relevant features for representing the data, resulting in a more objective process.
- *Learning relevant spatial relationships from neuroimaging data.* In the context of neuroimaging, a striking advantage of DL is its capacity to learn relevant spatial relationships among the image domain, such as an atrophy distributed across a

network of several brain regions supporting a specific function [89]. In classical ML techniques, the feature engineering step and the learning phase are dissociated, such that relevant spatial relationships may be lost. On the contrary, this spatial relationship might be preserved by DL techniques and integrated into the optimal hierarchy of features.

- *Learning non-linear relationships and biologically relevant compact representations.* As already discussed in section 2.5, non-linear relationships across data or dimensions relevant to NDDs are expected. Conceptually, the combination of the multiple layers available in DL architectures enables to encode this nonlinearity into a cascade of nonlinear transformations while reducing the input space into a lower dimensional 'latent space', providing a compact representation of the data. The recent works from [89]–[91] demonstrated that DL can exploit the presence of nonlinearity in neuroimaging data to learn generalizable representations highly relevant for characterizing the human brain. They combined supervised and unsupervised tasks in a DL framework which consisted in learning the representation from classification tasks (predicting age and sex) and then applying decomposition and clustering techniques to the latent space. These studies strongly support that DL approaches can provide more accurate mappings of the effects of age and sex on brain MRI than simpler models. The resulting representations obtained in these works are instrumental for refining the link between cognition and underlying brain systems. Another promising avenue of research denoted as Scientific Machine Learning (https://sciml.ai) consists in injecting traditional scientific mechanistic models into modern deep learning architectures in order to combine the benefits of efficient data-driven automatic learning with better interpretability and integration of biophysical constraints. See [92] for a review discussing the potential of these approaches in computational neuroscience and [93] for an example application to neuroimaging data. DL techniques can thus learn representations of data that have the potential to help explain the biological underpinnings of mental disorders, providing that enough data is available.

# 3. A non-exhaustive survey of existing papers on machine learning for NDDs and their limitations

We refer to the recent reviews [31], [32], [34], [51], [55], [69], [71], for a complete overview of the literature of the field. Here, we survey a selection of very recent works that we consider particularly relevant with respect to the opportunities offered by recent ML techniques applied to large open datasets, or that illustrate the challenges faced by current approaches, to be addressed in the near future.

## 3.1. Using ML techniques on neuroimaging data to predict the diagnosis

An international challenge (146 challengers) has been organized to predict ASD diagnosis based on several neuroimaging modalities [43]. This challenge was conducted on the largest sample available to date (>2000 individuals from the ABIDE dataset and a second, private dataset not open to challengers). An additional dataset from the EU-AIMS project [22] was used to evaluate the reproducibility of the prediction on an independent dataset (out-of-sample prediction). The ten best submissions either used logistic regression as a first layer predictor, linear vector classification or a combination of different methods. Best algorithms managed to predict ASD diagnosis with an in-sample AUC of 0.80. Resting-state fMRI data was a better diagnostic predictor than anatomical MRI, and simple logistic regression performed better than complex graph convolutional deep-learning models (likely due to overfitting). Finally, the performances of the best algorithms decreased to an out-of-sample AUC of 0.72 (on the external sample). Authors projected that 10,000 individuals might be necessary to reach the optimal prediction.

Another study of interest was led by the consortium ‘Infant Brain Imaging Study’-IBIS [94]. The authors investigated whether infants at high familial risk for autism present early postnatal atypical brain volume. A deep learning algorithm used surface area at 6 and 12 months to successfully predict an early diagnosis of autism in infants at high risk of autism at 24 months (in-sample predictive value of 81%, no out-of-sample prediction accuracy provided). These results should be tempered by several major pitfalls. First the diagnosis of

ASD is very challenging at that early age. Second, the sample size was very small (15 high-risk infants diagnosed with autism at 24 months) and thus does not comply with the recommended practices for predictive modeling [46]. Third, the specificity of the results with respect to other NDDs was not assessed. A confirmation of the reproducibility of these results in a larger, external cohort would thus be much welcome.

Overall, these results showed that applying prediction algorithms on large enough imaging data could be instrumental for early detection of ASD and therefore early intervention. In line with the conclusions of previous reviews [31], [69], these studies also demonstrated the relevance of using imaging data as an intermediate phenotype between the biological cause (e.g. deletion of the gene content at the 16p11.2 chromosomal segment) and the associated phenotype (e.g. ASD, ADHD, Intellectual Disability).

## 3.2. Latent space decomposition and subtyping approaches applied to NDDs

Complementary works are aiming to face clinical and biological heterogeneity in NDDs using a subtyping approach based on imaging data. Using hierarchical clustering methods on neuroanatomical data, Hong and colleagues [95] identified three distinct morphometric subtypes in ASD: ASD-I characterized by cortical thickening, increased surface area, tissue blurring; ASD-II with cortical thinning, decreased geodesic distance; and ASD-III with increased geodesic distance. These groups were associated with gradual symptom severities and might help tackle the well-known clinical heterogeneity issue introduced in section 1. The genetic contribution to the observed clinical heterogeneity was investigated across 8 psychiatric conditions including ASD and ADHD [96] with common variants. Exploratory factor analysis (EFA) on GWAS cross-disorders summary results led to identification of three genetically inter-related groups of disorders, explaining together 51% of the genetic variation across NDDs and psychiatric conditions. The first factor linked anorexia nervosa, OCD, and Tourette syndrome. The second one was associated with major depression, bipolar disorder, and schizophrenia. The last one encompassed early-onset NDDs (ASD, ADHD, Tourette syndrome) and major depression. Similar to EFA results, hierarchical genetic clustering identified the same three sub-groups among the eight disorders. These methods therefore have a great potential to uncover new biologically-relevant diagnostic categories.

Such overlaps across clinical diagnoses have also been characterized at the imaging level. [19] determined a common pattern of group differences in cortical thickness across 6 disorders –including ASD, OCD, ADHD, schizophrenia, bipolar, major depression disorders– and their link with gene expression profiles. Analyses of correlation and clustering revealed a shared profile of differences across disorders with 48% of variance explained, associated with pyramidal-cell gene expression. Analyses of gene co-expression highlighted two pre- and post-natal clusters associated with this common brain profile of group differences, enriched with genes associated with these disorders. Kebets and colleagues [97] applied partial least square regression (PLSR) to resting-state fMRI and cognitive metrics in participants with either ASD, ADHD, schizophrenia, or bipolar disorders. They identified three latents components (general psychopathology, cognitive dysfunction, and impulsivity) with unique fMRI signatures. Connectivity patterns of the somatosensory-motor network were main drivers across the 3 components. Similar findings on the somatosensory-motor network have been observed by [98], and extended to rare genetic mutations that confer high risk for neuropsychiatric conditions. [42] designed a hierarchical Bayesian modeling framework to derive hidden disease dimensions from RS-fMRI data across a population of ADHD, ASD, and controls. Using these methods, the number of components is inferred from the data. They obtained 45 hidden components that were then reduced to three main factors for better interpretation. For each of these three identified factors, the authors characterized the associated fMRI coupling patterns and symptom measures from the clinical questionnaires. These brain-derived factors predicted the classification of subjects as ADHD, ASD or control with an accuracy of 67%, computed using a variant of cross-validation called pre-validation described in [99]. This variant is expected to enable a fairer evaluation of the group labels than cross-validation, but still leaves room for errors compared to out-of-sample predictions [46].

Latent space decomposition techniques have been also used to identify general principles of the hierarchical brain organization –denoted as functional gradients– that locate sensory-motor networks at one end, and the transmodal default-mode network at the other end [100], [101]. Hong and colleagues [102] hypothesized that NDDs conditions may preferentially affect the sensory-motor dimension. They used surface-based analytical models to compare the first functional gradient (explaining 24% of the connectome variance) in ASD vs. controls and showed that both extremes of the rostrocaudal gradient were decreased in ASD. Interestingly, vertex-wise analyses revealed that such diminution in ASD was driven by transmodal medial PFC and posterior cingulate regions [102].

Combining large-scale multidimensional data is perceived as the golden standard to correctly apply ML algorithms. However, only a few precision medicine studies managed so far to do so. In [103], the authors extracted electronic health records, familial whole-exome sequences, and neurodevelopmental gene expression patterns in a large sample of ASD patients. Their goal was to identify biologically homogeneous ASD subtypes. For this purpose, the authors used spatiotemporal expression data from typically developing human brains to identify clusters of exons that are co-expressed during early human brain development. Based on prior knowledge on sexually-different prenatal gene expression in ASD, they focused the analysis on a set of clusters that are differentially expressed between males and females. They then selected inherited, likely gene-disrupting variants among all the ASD-segregating ones by leveraging a large dataset of families who have 1 child with ASD and 1 unaffected sibling. They mapped variants back to exon clusters to identify 33 clusters of neurodevelopmentally co-regulated, ASD-segregating deleterious variants. The functional enrichment analysis of the identified exon clusters (detailed in [103]) revealed a new molecular convergence on lipid regulation, with variants expected to collectively alter LDL, cholesterol and triglycerides levels. They confirmed that children with ASD have blood lipid profiles that are significantly outside the physiological range. Finally, they characterized the diagnostic spectrum of the dyslipidemia-associated ASD subtype and confirmed its specificity by comparing with individuals with ASD and no dyslipidemia. This work demonstrated the potential of combining massive amounts of multimodal data for uncovering new ASD subtypes.

## 3.3. Normative modeling

In [104], the authors applied normative modeling to a large sample of ASD and controls males covering a wide age range (5–40 years). They investigated the potential of age-related effects on cortical thickness to serve as an individualized metric of atypicality in individuals with ASD. They reported that only a small subgroup of patients showed age-atypical cortical thickness. By comparing with conventional case-control analyses, they observed that most case-control differences were driven by a small subgroup of patients with high atypicality for their age. Highly consistent results were obtained in another application of normative modeling to a different ASD cohort [105], despite important variations across these studies. The population of the second work was composed of both males and females and sex was included as a factor in the normative model. In addition, the normative models were estimated using different approaches (non-parametric regression in [104], Gaussian

process regression in [105]). The overall consistent results despite the methodological differences support the relevance of the normative modeling approach for NDDs. In a follow-up study, [106] applied the spectral clustering technique to the atypicality maps computed at the individual level as deviation in the cortical thickness with respect to the normative model estimated in [105].They identified five subtypes of individuals with ASD and assessed their separability using a multi-class linear SVM. Each subpopulation was then characterized in terms of demographic and clinical measures as well as association with polygenic scores for seven traits (autism, ADHD, epilepsy, Full IQ, neuroticism, schizophrenia, and cross disorder risk for psychiatric disorders). Importantly, they observed striking differences in the spatial patterns of cortical thickness atypicality maps between subtypes: 3 clusters showed reduced cortical thickness relative to the normative pattern whereas 2 clusters showed an increased cortical thickness. These distinct and opposing atypicalities across different subtypes could explain the inconsistency in the previous case–control analyses. A last study did apply normative modeling to an adult population of ADHD patients [107]. The authors estimated a normative model predicting regional gray and white matter volumes across the brain from age and sex. They observed deviations shared across patients in gray matter in the cerebellum, temporal regions, and the hippocampus. They also provided a measure of the inter-individual variation between ADHD patients with extreme deviations in specific regions in more than 2% of the participants. Overall, these results highlighted the relevance of the normative modeling approach to understanding the heterogeneity in NDDs.

## 3.4. Genetic features to predict cognitive deficit in NDDs.

As extensively discussed in [39], attempts to dissect mechanisms of NDD have mainly used a top-down approach, starting with a diagnosis and moving down to brain intermediate phenotypes and to genes. By contrast, the recruitment of groups based on the presence of a genetic risk factor for NDDs allows for the investigation of pathways related to a particular biological risk for psychiatric symptoms (bottom-up approach). Clinical routine with genomic microarrays revealed that copy number variants are present in 10 to 15% of children with neurodevelopmental conditions [108]. Genetic-first approaches can however only be applied to a few recurrent pathogenic mutations frequent enough to establish a case-control study design. Thus, the effect of the vast majority of rare deleterious risk variants remains undocumented. Because a highly diverse landscape of rare variants confers a higher risk to a spectrum of NDDs, studies focusing on individual mutations will not be able to properly

disentangle the relationship between mutations, molecular mechanisms, and diagnoses. Huguet and colleagues [109] speculated that large effect size pathogenic deletions may be attributable to the sum of individual effects of genes encompassed in each copy number variation. They introduced a new framework to estimate the effect of any pathogenic deletion on intelligence quotient (IQ). Using several types of functional annotations of rare genetic deletions associated with NDDs, the proposed framework predicted their impact on IQ with 76% accuracy [109]. They showed that haploinsufficiency scores –probability of being loss of function intolerant (pLI)– best explain the cognitive deficits. Follow up works specifically on ASD confirmed that this score was the best predictor of IQ deficit and autism risk (odd-ratio) [110], [111]. Deletion of 1 point of pLI was associated with a decrease of 2.6 points of IQ in autism.

## 3.5. Deep learning applied to NDDs

A deep-learning based framework has been recently introduced to predict the regulatory contribution of non-coding mutations to autism [112]. Authors constructed a deep convolutional network to model the functional impact of each individual mutation (single nucleotide polymorphism). They first identified that ASD probands (n=1700 families) were carriers of a higher rate of transcriptional and post-transcriptional regulation disrupting de novo mutations compared with their siblings. They also revealed a convergent pattern of coding and non-coding mutations.

In [113], the authors analyzed Resting State-fMRI (RS-fMRI) data from 260 subjects with ADHD and 343 healthy controls from the ADHD-200 database. They proposed to represent RS-fMRI data from each individual as a graph that integrates both temporal and spatial correlation of regional time-series signals. An original graph convolutional neural network architecture was introduced to characterize the brain functional connectome. The model also included seven non-imaging variables (age, gender, handedness, IQ measurement, and three Wechsler Intelligence Scale evaluation IQ variables) and was trained to distinguish ADHD patients from HC. Several experiments showed a performance gain compared to previous methods including SVM, logistic regression and conventional graph convolutional networks. The proposed method outperformed other competing approaches, including SVM and logistic regression, with an AUC of 75 (72.0% accuracy, 71.6% specificity, and 72.2% sensitivity) on a 10-fold cross-validation. A leave-study-site-out experiment demonstrated the robustness of the proposed model for unseen data from different study sites, and

experiments with simplified versions of the model showed the relevance of each proposed improvement. Most discriminative regions were mainly located in the frontal lobe, occipital lobe, subcortical lobe, temporal lobe, and cerebellum –with hypo-connections mainly between the frontal, parietal, and temporal lobes and widespread hyper-connections.

These studies support that new methodological improvements can be expected from the very active field of deep-learning applications to neuroimaging and genetics data. As pointed in [88], the anticipated increase in sample size in NDDs studies will allow fit more complex models, which might reveal larger differences in performances compared to conventional methods. The literature of DL applications to NDDs is however still in its initial stages and major challenges such as tendency to over-fitting [43] have to be carefully addressed in future studies.

## 3.6. Discussion

The review of the selected recent studies presented above demonstrates that applications of ML in NDDs is a very active field of research, with encouraging perspectives. This field indeed benefits directly from initiatives to openly share data [114], which did increase the sample size involved across studies, and favored the engagement of ML scientists. The paradigm shift from diagnostic-first to genetic-first and from one diagnostic at a time to cross-diagnoses approaches is afoot, with a clear rise of large scale studies based on normative modeling and deep-learning approaches. Methodological works continue to introduce new innovative ML approaches specifically designed to address the central tasks in NDDs. Importantly, the adoption of best practices for validation and replication of the results across independent datasets as stated in [46] is clearly encouraged by the recent reviews [31], [32], [34], [51], [55], [69], [71]. However, the validation is limited by insufficient access to large enough datasets combining multiscale data (genetics, transcriptomic, proteomic, metabolomic, neuroimaging features, phenomics). There is no open dataset so far offering that level of granularity. Indeed, the imaging field is just reaching the sample size allowing for running modern ML techniques for some but not all modalities. For instance, large scale studies involving Diffusion Weighted Imaging are clearly lacking in NDDs, probably due to insufficient access to appropriate data. The genomic field is not ready yet, and several domains remain relatively new (e.g., first genome sequenced in 2000', Next-Generation Sequencing techniques in 2010') and expensive (e.g., RNA-Seq data)

[115], [116]. Such data will provide –in the near future– massive potential for accurate classification and appropriate validation.

# 4. Open challenges and conclusion

Methodological improvements described in section 2 and studies reviewed in section 3 are encouraging for concrete impact on clinical practice in the future. However, such clinical translation is raising major challenges that should be addressed.

## 4.1. Potential bias in data and processing pipelines

Despite the large amount of new approaches released by recent literature, some potential biases in analysis pipelines should be mentioned. For instance, the analysis of functional networks computed from RS-fMRI rely on a complex succession of processing steps. Several of these processing steps actually correspond to implementing assumptions regarding the data. However, the validity of these assumptions and their influence on the subsequent results are not sufficiently discussed in the literature. See for instance [117] for a quantitative evaluation of the impact of the brain parcellation procedure on functional connectivity analyses. Another major barrier to reproducibility is the lack of compatibility among programming languages, software versions and operating systems as illustrated in [118]. This report highlights the challenges and potential solutions to be implemented both at the individual researcher and community levels in order to enable the appropriate reuse of published methods.

On the data side, the limitations related to the absence of recording of potentially influencing factors are not sufficiently investigated and acknowledged. As pointed e.g. in [119]: “The extent of brain differences in disease may depend critically on a patient's age, duration of illness, course of treatment, as well as adherence to the treatment, polypharmacy and other unmeasured factors. Differences in ancestral background, as determined based on genotype, are strongly related to systematic differences in brain shape. Any realistic understanding of the brain imaging measures must take all these into account, as well as acknowledge the existence of causal factors perhaps not yet known or even imagined.” As a concrete illustration, [120], [121] recently reported significant alterations in brain

morphometry induced by prematurity, a factor that was not considered by any of the studies we reviewed here. Such uncontrolled factors might introduce considerable bias in the learning process. The ML research field has identified this pitfall and several solutions to prevent unexpected implications in clinical applications are actively debated [122]–[124].

## 4.2. Interpretability and biological substrates

Even in the absence of bias, the interpretation of the outcome of any ML algorithm in the context of clinical application represents a critical challenge. More than the level of raw performance, the level of expertise required from medical doctors in 1) the recording and 2) the analysis of the data compared to 'expertise-free' raw data is a question that requires more attention. We refer to [125] for a thoughtful discussion on the need for clarification of the role of ML-based tools in relation to clinicians decisions and actions in clinical practice. The authors call for a more systematic demonstration that models learning from non-clinician-initiated data outperform models based on clinician-initiated data. They purposely argue that models driven by features derived from the actions of clinicians and not related to the underlying physiology might introduce some deleterious circularity. Indeed, the outcome of such a model might potentially confuse more than support a clinician in his decisions.

Then –regarding the interpretation in terms of pathophysiology– the challenge is to relate the decisions of any ML techniques to putative underlying biological processes. Methodological innovations will enhance the explainability of ML models, but explainability and transparency do not imply interpretability [126], [127]. Another major challenge is to assess the biological relevance of the features extracted from the data during the learning procedure. Purely data driven approaches are limited by the difficulty to relate the parameters of the model to biological knowledge. A promising perspective consists in inserting biological priors directly in predictive models. See [92] for an introductory review to this type of approach in the context of computational neuroscience and (https://sciml.ai) for further information on the emerging field of Scientific Machine Learning. However, extensive basic research at conceptual, methodological and experimental levels are required to fill the gap between measures accessible in-vivo in patients and the biophysiology acting at cellular- and molecular-levels. See for instance [128] for an illustration of the complexity of this challenge, where the authors propose a framework integrating different levels of interactions, from genes to cells, circuits, and clinical expression, to better understand and treat cortical

malformations. As discussed in [129] for ASD, research designs aiming at a better conceptual integration between different levels of brain organization are required to characterize the cascade of pathogenic processes in NDDs.

### 4.3. Conclusion

In NDDs, as in healthcare in general, ML has a role to play in addressing the longstanding deficiencies such as serious diagnostic errors, mistakes in treatment, and waste of resources [130]. Indeed, ML will undoubtedly help redefine NDDs categories and other mental illnesses more objectively, identify them at an early stage, and contribute to more adapted treatments. The rise of ML is the occasion to improve the standardization of practice and to enforce the generalization of open science with preregistration and data sharing or federated learning. In addition, the field has to demonstrate high and reproducible performances in the real-world clinical environment. Finally, major conceptual, ethical and socio-technical challenges have to be addressed.

# Acknowledgements

We would like to thank the editor and Guillaume Dumas who served as a reviewer for their insightful feedback that has improved our manuscript. This work was supported by the French government under management of Agence Nationale de la Recherche, reference ANR-19-CE45-0014

# References


[1] American Psychiatric Association, *The Diagnostic and Statistical Manual of Mental Disorders: DSM 5*. Arlington, VA: American Psychiatric Publishing, 2013.

[2] S. Jacob, J. J. Wolff, M. S. Steinbach, C. B. Doyle, V. Kumar, and J. T. Elison, "Neurodevelopmental heterogeneity and computational approaches for understanding autism," *Transl. Psychiatry*, vol. 9, no. 1, pp. 1–12, Feb. 2019, doi: 10.1038/s41398-019-0390-0.

[3] M. V. Lombardo, M.-C. Lai, and S. Baron-Cohen, "Big data approaches to decomposing heterogeneity across the autism spectrum," *Mol. Psychiatry*, vol. 24, no. 10, Art. no. 10, Oct. 2019, doi: 10.1038/s41380-018-0321-0.

[4] S. E. Hyman, "Can neuroscience be integrated into the DSM-V?," *Nat. Rev. Neurosci.*, vol. 8, no. 9, pp. 725–732, Sep. 2007, doi: 10.1038/nrn2218.

[5] T. Bourgeron, "What Do We Know about Early Onset Neurodevelopmental Disorders?," Aug. 2015, doi: 10.7551/mitpress/9780262029865.003.0005.

[6] G. Joshi *et al.*, "The Heavy Burden of Psychiatric Comorbidity in Youth with Autism Spectrum Disorders: A Large Comparative Study of a Psychiatrically Referred Population," *J. Autism Dev. Disord.*, vol. 40, no. 11, pp. 1361–1370, Nov. 2010, doi: 10.1007/s10803-010-0996-9.

[7] M.-C. Lai, M. V. Lombardo, and S. Baron-Cohen, "Autism," *The Lancet*, vol. 383, no. 9920, pp. 896–910, Mar. 2014, doi: 10.1016/S0140-6736(13)61539-1.

[8] V. Anttila *et al.*, "Analysis of shared heritability in common disorders of the brain," *Science*, vol. 360, no. 6395, p. eaap8757, Jun. 2018, doi: 10.1126/science.aap8757.

[9] R. Siugzdaite, J. Bathelt, J. Holmes, and D. E. Astle, "Transdiagnostic Brain Mapping in Developmental Disorders," *Curr. Biol.*, vol. 30, no. 7, pp. 1245-1257.e4, Apr. 2020, doi: 10.1016/j.cub.2020.01.078.

[10] C. S. Leblond *et al.*, "Operative list of genes associated with autism and neurodevelopmental disorders based on database review," *Mol. Cell. Neurosci.*, vol. 113, p. 103623, Jun. 2021, doi: 10.1016/j.mcn.2021.103623.

[11] K. L. Mills *et al.*, "Inter-individual variability in structural brain development from late childhood to young adulthood," *NeuroImage*, vol. 242, p. 118450, Nov. 2021, doi: 10.1016/j.neuroimage.2021.118450.

[12] T. T. Brown, "Individual differences in human brain development," *Wiley Interdiscip. Rev. Cogn. Sci.*, vol. 8, no. 1–2, p. e1389, Jan. 2017, doi: 10.1002/wcs.1389.

[13] T. T. Brown *et al.*, "Neuroanatomical Assessment of Biological Maturity," *Curr. Biol.*, vol. 22, no. 18, pp. 1693–1698, Sep. 2012, doi: 10.1016/j.cub.2012.07.002.

[14] D. K. Thompson *et al.*, "Tracking regional brain growth up to age 13 in children born term and very preterm," *Nat. Commun.*, vol. 11, no. 1, p. 696, Dec. 2020, doi: 10.1038/s41467-020-14334-9.

[15] D. Witvliet *et al.*, "Connectomes across development reveal principles of brain maturation," *Nature*, vol. 596, no. 7871, pp. 257–261, Aug. 2021, doi: 10.1038/s41586-021-03778-8.

[16] A. M. Fjell *et al.*, "Continuity and Discontinuity in Human Cortical Development and Change From Embryonic Stages to Old Age," *Cereb. Cortex*, vol. 29, no. 9, pp. 3879–3890, Aug. 2019, doi: 10.1093/cercor/bhy266.

[17] R. K. Reh *et al.*, "Critical period regulation across multiple timescales," *Proc. Natl. Acad. Sci.*, vol. 117, no. 38, pp. 23242–23251, Sep. 2020, doi: 10.1073/pnas.1820836117.

[18] R. G. Rudel, "Residual Effects of Childhood Reading Disabilities," *Bull. Orton Soc.*,

vol. 31, pp. 89–102, 1981.
[19] Y. Patel *et al.*, “Virtual Histology of Cortical Thickness and Shared Neurobiology in 6 Psychiatric Disorders,” *JAMA Psychiatry*, 2020, doi: 10.1001/jamapsychiatry.2020.2694.
[20] B. J. Casey *et al.*, “The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites,” *Dev. Cogn. Neurosci.*, vol. 32, no. January, pp. 43–54, Aug. 2018, doi: 10.1016/j.dcn.2018.03.001.
[21] a Di Martino *et al.*, “The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism.,” *Mol. Psychiatry*, vol. 19, no. April, pp. 659–67, 2014, doi: 10.1038/mp.2013.78.
[22] E. Loth *et al.*, “The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders,” *Mol. Autism*, vol. 8, no. 1, p. 24, Dec. 2017, doi: 10.1186/s13229-017-0146-8.
[23] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, “The Neuro Bureau ADHD-200 Preprocessed repository,” *NeuroImage*, vol. 144, pp. 275–286, Jan. 2017, doi: 10.1016/j.neuroimage.2016.06.034.
[24] T. May *et al.*, “Trends in the Overlap of Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder: Prevalence, Clinical Management, Language and Genetics,” *Curr. Dev. Disord. Rep.*, vol. 5, no. 1, pp. 49–57, Mar. 2018, doi: 10.1007/s40474-018-0131-8.
[25] R. Mansour, A. T. Dovi, D. M. Lane, K. A. Loveland, and D. A. Pearson, “ADHD severity as it relates to comorbid psychiatric symptomatology in children with Autism Spectrum Disorders (ASD),” *Res. Dev. Disabil.*, vol. 60, pp. 52–64, Jan. 2017, doi: 10.1016/j.ridd.2016.11.009.
[26] M. Hoogman *et al.*, “Consortium neuroscience of attention deficit/hyperactivity disorder and autism spectrum disorder: The ENIGMA adventure,” *Hum. Brain Mapp.*, vol. n/a, no. n/a, 2021, doi: https://doi.org/10.1002/hbm.25029.
[27] E. Fombonne, “The epidemiology of autism: a review,” *Psychol. Med.*, vol. 29, no. 4, pp. 769–786, Jul. 1999, doi: 10.1017/S0033291799008508.
[28] T. Bourgeron, “From the genetic architecture to synaptic plasticity in autism spectrum disorder,” *Nat. Rev. Neurosci.*, vol. 16, no. 9, pp. 551–563, Sep. 2015, doi: 10.1038/nrn3992.
[29] C. Lord *et al.*, “Autism spectrum disorder,” *Nat. Rev. Dis. Primer*, vol. 6, no. 1, Art. no. 1, Jan. 2020, doi: 10.1038/s41572-019-0138-4.
[30] A. Thapar and M. Cooper, “Attention deficit hyperactivity disorder,” *The Lancet*, vol. 387, no. 10024, pp. 1240–1250, Mar. 2016, doi: 10.1016/S0140-6736(15)00238-X.
[31] T. Wolfers *et al.*, “From pattern classification to stratification: towards conceptualizing the heterogeneity of Autism Spectrum Disorder,” *Neurosci. Biobehav. Rev.*, vol. 104, no. April, pp. 240–254, Sep. 2019, doi: 10.1016/j.neubiorev.2019.07.010.
[32] M. Xu, V. Calhoun, R. Jiang, W. Yan, and J. Sui, “Brain imaging-based machine learning in autism spectrum disorder: methods and applications,” *J. Neurosci. Methods*, vol. 361, p. 109271, Sep. 2021, doi: 10.1016/j.jneumeth.2021.109271.
[33] C. S. Hiremath *et al.*, “Emerging behavioral and neuroimaging biomarkers for early and accurate characterization of autism spectrum disorders: a systematic review,” *Transl. Psychiatry*, vol. 11, no. 1, pp. 1–12, Jan. 2021, doi: 10.1038/s41398-020-01178-6.
[34] D. Bzdok and A. Meyer-Lindenberg, “Machine Learning for Precision Psychiatry: Opportunities and Challenges,” *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 3, no. 3, pp. 223–230, Mar. 2018, doi: 10.1016/j.bpsc.2017.11.007.
[35] K. K. Hyde *et al.*, “Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review,” *Rev. J. Autism Dev. Disord.*, vol. 6, no. 2, pp. 128–146,

Jun. 2019, doi: 10.1007/s40489-019-00158-x.
[36] T. Eslami, F. Almuqhim, J. S. Raiker, and F. Saeed, “Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention- Deficit/Hyperactivity Disorder Using Functional and Structural MRI: A Survey,” *Front. Neuroinformatics*, vol. 14, 2021, Accessed: Jan. 21, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2020.575999
[37] L. Mottron and D. Bzdok, “Autism spectrum heterogeneity: fact or artifact?,” *Mol. Psychiatry*, vol. 25, no. 12, pp. 3178–3185, Dec. 2020, doi: 10.1038/s41380-020-0748-y.
[38] E. Loth *et al.*, “The meaning of significant mean group differences for biomarker discovery,” *PLOS Comput. Biol.*, vol. 17, no. 11, p. e1009477, Nov. 2021, doi: 10.1371/journal.pcbi.1009477.
[39] C. A. Moreau, A. Raznahan, P. Bellec, M. Chakravarty, P. M. Thompson, and S. Jacquemont, “Dissecting autism and schizophrenia through neuroimaging genomics,” *Brain*, no. awab096, Mar. 2021, doi: 10.1093/brain/awab096.
[40] J. P. Simmons, L. D. Nelson, and U. Simonsohn, “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant,” *Psychol. Sci.*, vol. 22, no. 11, pp. 1359–1366, Nov. 2011, doi: 10.1177/0956797611417632.
[41] A. F. Marquand, S. M. Kia, M. Zabihi, T. Wolfers, J. K. Buitelaar, and C. F. Beckmann, “Conceptualizing mental disorders as deviations from normative functioning,” *Mol. Psychiatry*, Jun. 2019, doi: 10.1038/s41380-019-0441-1.
[42] J. M. Kernbach *et al.*, “Shared endo-phenotypes of default mode dysfunction in attention deficit/hyperactivity disorder and autism spectrum disorder,” *Transl. Psychiatry*, vol. 8, no. 1, pp. 1–11, Jul. 2018, doi: 10.1038/s41398-018-0179-6.
[43] N. Traut *et al.*, “Insights from an autism imaging biomarker challenge: Promises and threats to biomarker discovery,” *NeuroImage*, vol. 255, p. 119171, Jul. 2022, doi: 10.1016/j.neuroimage.2022.119171.
[44] L. Maier-Hein *et al.*, “Why rankings of biomedical image analysis competitions should be interpreted with care,” *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Dec. 2018, doi: 10.1038/s41467-018-07619-7.
[45] E. E. Bron *et al.*, “Ten years of image analysis and machine learning competitions in dementia,” *ArXiv211207922 Cs*, Dec. 2021, Accessed: Dec. 21, 2021. [Online]. Available: http://arxiv.org/abs/2112.07922
[46] R. A. Poldrack, G. Huckins, and G. Varoquaux, “Establishment of Best Practices for Evidence for Prediction: A Review,” *JAMA Psychiatry*, vol. 77, no. 5, pp. 534–540, May 2020, doi: 10.1001/jamapsychiatry.2019.3671.
[47] G. Varoquaux, “Cross-validation failure: Small sample sizes lead to large error bars,” *NeuroImage*, vol. 180, pp. 68–77, Oct. 2018, doi: 10.1016/j.neuroimage.2017.06.061.
[48] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion, “Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines,” *NeuroImage*, vol. 145, pp. 166–179, 2017, doi: 10.1016/j.neuroimage.2016.10.038.
[49] X. Bouthillier *et al.*, “Accounting for Variance in Machine Learning Benchmarks,” *ArXiv210303098 Cs Stat*, Mar. 2021, Accessed: Dec. 21, 2021. [Online]. Available: http://arxiv.org/abs/2103.03098
[50] P. Kassraian-Fard, C. Matthis, J. H. Balsters, M. H. Maathuis, and N. Wenderoth, “Promises, Pitfalls, and Basic Guidelines for Applying Machine Learning Classifiers to Psychiatric Imaging Data, with Autism as an Example,” *Front. Psychiatry*, vol. 7, p. 177, 2016, doi: 10.3389/fpsyt.2016.00177.
[51] D. Bzdok and J. P. A. Ioannidis, “Exploration, Inference, and Prediction in Neuroscience and Biomedicine,” *Trends Neurosci.*, vol. 42, no. 4, pp. 251–262, Apr.

2019, doi: 10.1016/j.tins.2019.02.001.
[52] A. Abraham *et al.*, “Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example,” *NeuroImage*, vol. 147, no. October 2016, pp. 736–745, Feb. 2017, doi: 10.1016/j.neuroimage.2016.10.045.
[53] H. Johansen-Berg and T. E. J. Behrens, *Diffusion MRI*, Academic p. Elsevier, 2014. doi: 10.1016/C2011-0-07047-3.
[54] S. W. Choi, T. S.-H. Mak, and P. F. O’Reilly, “Tutorial: a guide to performing polygenic risk score analyses,” *Nat. Protoc.*, vol. 15, no. 9, pp. 2759–2772, Sep. 2020, doi: 10.1038/s41596-020-0353-1.
[55] L. Q. Uddin, D. R. Dajani, W. Voorhies, H. Bednarz, and R. K. Kana, “Progress and roadblocks in the search for brain-based biomarkers of autism and attention-deficit/hyperactivity disorder,” *Transl. Psychiatry*, vol. 7, no. 8, Art. no. 8, Aug. 2017, doi: 10.1038/tp.2017.164.
[56] C. Cameron *et al.*, “The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives,” *Front. Neuroinformatics*, vol. 7, 2013, doi: 10.3389/conf.fninf.2013.09.00041.
[57] E.-M. Rødgaard, K. Jensen, J.-N. Vergnes, I. Soulières, and L. Mottron, “Temporal Changes in Effect Sizes of Studies Comparing Individuals With and Without Autism: A Meta-analysis,” *JAMA Psychiatry*, vol. 76, no. 11, pp. 1124–1132, Nov. 2019, doi: 10.1001/jamapsychiatry.2019.1956.
[58] T. Insel *et al.*, “Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders,” *Am. J. Psychiatry*, vol. 167, no. 7, pp. 748–751, Jul. 2010, doi: 10.1176/appi.ajp.2010.09091379.
[59] T. R. Insel and B. N. Cuthbert, “Brain disorders? Precisely,” *Science*, vol. 348, no. 6234, pp. 499–500, May 2015, doi: 10.1126/science.aab2358.
[60] B. N. Cuthbert and T. R. Insel, “Toward the future of psychiatric diagnosis: the seven pillars of RDoC,” *BMC Med.*, vol. 11, no. 1, p. 126, May 2013, doi: 10.1186/1741-7015-11-126.
[61] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, “The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians,” *J. Autism Dev. Disord.*, vol. 31, no. 1, pp. 5–17, Feb. 2001, doi: 10.1023/A:1005653411471.
[62] J. N. Constantino and C. P. Gruber, *Social responsiveness scale: SRS-2*. Western Psychological Services Torrance, CA, 2012.
[63] A. Ronald and R. A. Hoekstra, “Autism spectrum disorders and autistic traits: A decade of new twin studies,” *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, vol. 156, no. 3, pp. 255–274, 2011, doi: 10.1002/ajmg.b.31159.
[64] E. Feczko, O. Miranda-Dominguez, M. Marr, A. M. Graham, J. T. Nigg, and D. A. Fair, “The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes,” *Trends Cogn. Sci.*, vol. 23, no. 7, pp. 584–601, Jul. 2019, doi: 10.1016/j.tics.2019.03.009.
[65] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. New York, NY: Springer, 2013. doi: 10.1007/978-1-4614-7138-7_1.
[66] S. Tang, N. Sun, D. L. Floris, X. Zhang, A. D. Martino, and B. T. T. Yeo, “Reconciling Dimensional and Categorical Models of Autism Heterogeneity: A Brain Connectomics and Behavioral Study,” *Biol. Psychiatry*, vol. 87, no. 12, pp. 1071–1082, Jun. 2020, doi: 10.1016/j.biopsych.2019.11.009.
[67] E. Feczko and D. A. Fair, “Methods and Challenges for Assessing Heterogeneity,” *Biol. Psychiatry*, 2020, doi: 10.1016/j.biopsych.2020.02.015.
[68] U. von Luxburg, R. C. Williamson, and I. Guyon, “Clustering: Science or Art?,” in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, Jun. 2012, pp. 65–79. Accessed: Jan. 12, 2022. [Online]. Available: https://proceedings.mlr.press/v27/luxburg12a.html

[69] S.-J. Hong *et al.*, “Toward Neurosubtypes in Autism,” *Biol. Psychiatry*, vol. 88, no. 1, pp. 111–128, Jul. 2020, doi: 10.1016/j.biopsych.2020.03.022.
[70] C. W. Nordahl *et al.*, “The Autism Phenome Project: Toward Identifying Clinically Meaningful Subgroups of Autism,” *Front. Neurosci.*, vol. 15, 2022, Accessed: Jan. 21, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2021.786220
[71] J. A. Agelink van Rentergem, M. K. Deserno, and H. M. Geurts, “Validation strategies for subtypes in psychiatry: A systematic review of research on autism spectrum disorder,” *Clin. Psychol. Rev.*, vol. 87, p. 102033, Jul. 2021, doi: 10.1016/j.cpr.2021.102033.
[72] D. Bzdok and B. T. T. Yeo, “Inference in the age of big data: Future perspectives on neuroscience,” *NeuroImage*, vol. 155, no. April, pp. 549–564, Jul. 2017, doi: 10.1016/j.neuroimage.2017.04.061.
[73] A. D. Grotzinger *et al.*, “Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits,” *Nat. Hum. Behav.*, vol. 3, no. 5, Art. no. 5, May 2019, doi: 10.1038/s41562-019-0566-x.
[74] C. Modenato *et al.*, “Lessons learnt from neuroimaging studies of Copy Number Variants, a systematic review,” *Biol. Psychiatry*, p. S0006322321013949, Jun. 2021, doi: 10.1016/j.biopsych.2021.05.028.
[75] A. F. Marquand, I. Rezek, J. Buitelaar, and C. F. Beckmann, “Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies,” *Biol. Psychiatry*, vol. 80, no. 7, pp. 552–561, Oct. 2016, doi: 10.1016/j.biopsych.2015.12.023.
[76] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.
[77] B. Xu, R. Kuplicki, S. Sen, and M. P. Paulus, “The pitfalls of using Gaussian Process Regression for normative modeling,” *PLOS ONE*, vol. 16, no. 9, p. e0252108, Sep. 2021, doi: 10.1371/journal.pone.0252108.
[78] A. Lefebvre *et al.*, “Alpha Waves as a Neuromarker of Autism Spectrum Disorder: The Challenge of Reproducibility and Heterogeneity,” *Front. Neurosci.*, vol. 12, 2018, Accessed: May 10, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2018.00662
[79] G. Chen *et al.*, “Beyond linearity in neuroimaging: Capturing nonlinear relationships with application to longitudinal studies,” *NeuroImage*, vol. 233, p. 117891, Jun. 2021, doi: 10.1016/j.neuroimage.2021.117891.
[80] C. J. Fraza, R. Dinga, C. F. Beckmann, and A. F. Marquand, “Warped Bayesian linear regression for normative modelling of big data,” *NeuroImage*, vol. 245, p. 118715, Dec. 2021, doi: 10.1016/j.neuroimage.2021.118715.
[81] A. M. Fjell *et al.*, “When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies.,” *NeuroImage*, vol. 50, no. 4, pp. 1376–83, May 2010, doi: 10.1016/j.neuroimage.2010.01.061.
[82] A. M. Fjell *et al.*, “Development and aging of cortical thickness correspond to genetic organization patterns,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 50, pp. 15462–15467, Dec. 2015, doi: 10.1073/pnas.1508831112.
[83] R. Dinga, C. J. Fraza, J. M. M. Bayer, S. M. Kia, C. F. Beckmann, and A. F. Marquand, “Normative modeling of neuroimaging data using generalized additive models of location scale and shape.” bioRxiv, p. 2021.06.14.448106, Jun. 14, 2021. doi: 10.1101/2021.06.14.448106.
[84] R. a. I. Bethlehem *et al.*, “Brain charts for the human lifespan,” *Nature*, vol. 604, no. 7906, Art. no. 7906, Apr. 2022, doi: 10.1038/s41586-022-04554-y.
[85] D. Batalle, A. D. Edwards, and J. O’Muircheartaigh, “Annual Research Review: Not just a small adult brain: understanding later neurodevelopment through imaging the

neonatal brain," *J. Child Psychol. Psychiatry*, vol. 59, no. 4, pp. 350–371, Apr. 2018, doi: 10.1111/jcpp.12838.
[86] S. Rutherford *et al.*, "The Normative Modeling Framework for Computational Psychiatry." bioRxiv, p. 2021.08.08.455583, Aug. 10, 2021. doi: 10.1101/2021.08.08.455583.
[87] S. Rutherford *et al.*, "Charting brain growth and aging at high spatial precision," *eLife*, vol. 11, p. e72904, Feb. 2022, doi: 10.7554/eLife.72904.
[88] S. Vieira, W. H. L. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neurosci. Biobehav. Rev.*, vol. 74, pp. 58–75, 2017, doi: 10.1016/j.neubiorev.2017.01.002.
[89] A. Abrol *et al.*, "Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning," *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, Jan. 2021, doi: 10.1038/s41467-020-20655-6.
[90] M. Zabihi *et al.*, "Non-linearity matters: a deep learning solution to generalization of hidden brain patterns across population cohorts," Oct. 2021. doi: 10.1101/2021.03.10.434856.
[91] W. H. L. Pinaya, A. Mechelli, and J. R. Sato, "Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study," *Hum. Brain Mapp.*, vol. 40, no. 3, pp. 944–954, 2019, doi: 10.1002/hbm.24423.
[92] M. R. Panahi, G. Abrevaya, J.-C. Gagnon-Audet, V. Voleti, I. Rish, and G. Dumas, "Generative Models of Brain Dynamics -- A review," *ArXiv211212147 Q-Bio*, Dec. 2021, Accessed: May 10, 2022. [Online]. Available: http://arxiv.org/abs/2112.12147
[93] G. Abrevaya *et al.*, "Learning Brain Dynamics With Coupled Low-Dimensional Nonlinear Oscillators and Deep Recurrent Networks," *Neural Comput.*, vol. 33, no. 8, pp. 2087–2127, Jul. 2021, doi: 10.1162/neco_a_01401.
[94] H. C. Hazlett *et al.*, "Early brain development in infants at high risk for autism spectrum disorder," *Nature*, vol. 542, no. 7641, pp. 348–351, 2017, doi: 10.1038/nature21369.
[95] S.-J. Hong, S. L. Valk, A. Di Martino, M. P. Milham, and B. C. Bernhardt, "Multidimensional Neuroanatomical Subtyping of Autism Spectrum Disorder," *Cereb. Cortex*, no. Betancur 2011, pp. 1–11, Sep. 2017, doi: 10.1093/cercor/bhx229.
[96] P. H. Lee *et al.*, "Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders," *Cell*, vol. 179, no. 7, pp. 1469-1482.e11, Dec. 2019, doi: 10.1016/j.cell.2019.11.020.
[97] V. Kebets *et al.*, "Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology," *Biol. Psychiatry*, vol. 86, no. 10, pp. 779–791, Nov. 2019, doi: 10.1016/j.biopsych.2019.06.013.
[98] C. A. Moreau *et al.*, "Mutations associated with neuropsychiatric conditions delineate functional brain connectivity dimensions contributing to autism and schizophrenia," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41467-020-18997-2.
[99] R. J. Tibshirani and B. Efron, "Pre-validation and inference in microarrays," *Stat. Appl. Genet. Mol. Biol.*, vol. 1, no. 1, Aug. 2002, doi: 10.2202/1544-6115.1000.
[100] J. M. Huntenburg, P.-L. Bazin, and D. S. Margulies, "Large-Scale Gradients in Human Cortical Organization," *Trends Cogn. Sci.*, vol. 22, no. 1, pp. 21–31, Jan. 2018, doi: 10.1016/j.tics.2017.11.002.
[101] D. S. Margulies *et al.*, "Situating the default-mode network along a principal gradient of macroscale cortical organization," *Proc. Natl. Acad. Sci.*, vol. 113, no. 44, pp. 12574–12579, Nov. 2016, doi: 10.1073/pnas.1608282113.
[102] S.-J. Hong *et al.*, "Atypical functional connectome hierarchy in autism," *Nat. Commun.*, vol. 10, no. 1, Art. no. 1, Mar. 2019, doi: 10.1038/s41467-019-08944-1.

[103] Y. Luo *et al.*, “A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia,” *Nat. Med.*, vol. 26, no. 9, Art. no. 9, Sep. 2020, doi: 10.1038/s41591-020-1007-0.
[104] R. a. I. Bethlehem, J. Seidlitz, R. Romero-Garcia, S. Trakoshis, G. Dumas, and M. V. Lombardo, “A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder,” *Commun. Biol.*, vol. 3, no. 1, p. 486, Dec. 2020, doi: 10.1038/s42003-020-01212-9.
[105] M. Zabihi *et al.*, “Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models,” *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 4, no. 6, pp. 567–578, Jun. 2019, doi: 10.1016/j.bpsc.2018.11.013.
[106] M. Zabihi *et al.*, “Fractionating autism based on neuroanatomical normative modeling,” *Transl. Psychiatry*, vol. 10, no. 1, p. 384, Dec. 2020, doi: 10.1038/s41398-020-01057-0.
[107] T. Wolfers, C. F. Beckmann, M. Hoogman, J. K. Buitelaar, B. Franke, and A. F. Marquand, “Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models,” *Psychol. Med.*, vol. 50, no. 2, pp. 314–323, Jan. 2020, doi: 10.1017/S0033291719000084.
[108] H. M. Kearney, E. C. Thorland, K. K. Brown, F. Quintero-Rivera, and S. T. South, “American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants,” *Genet. Med.*, vol. 13, no. 7, pp. 680–685, Jul. 2011, doi: 10.1097/GIM.0b013e3182217a3a.
[109] G. Huguet *et al.*, “Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples,” *JAMA Psychiatry*, vol. 75, no. 5, pp. 447–457, May 2018, doi: 10.1001/jamapsychiatry.2018.0039.
[110] E. Douard *et al.*, “Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome,” *Am. J. Psychiatry*, vol. 178, no. 1, pp. 87–98, Sep. 2020, doi: 10.1176/appi.ajp.2020.19080834.
[111] G. Huguet *et al.*, “Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability,” *Mol. Psychiatry*, vol. 26, no. 6, Art. no. 6, Jun. 2021, doi: 10.1038/s41380-020-00985-z.
[112] J. Zhou *et al.*, “Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk,” *Nat. Genet.*, vol. 51, no. 6, Art. no. 6, Jun. 2019, doi: 10.1038/s41588-019-0420-0.
[113] K. Zhao, B. Duka, H. Xie, D. J. Oathes, V. Calhoun, and Y. Zhang, “A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD,” *NeuroImage*, vol. 246, p. 118774, Feb. 2022, doi: 10.1016/j.neuroimage.2021.118774.
[114] M. P. Milham *et al.*, “Assessment of the impact of shared brain imaging data on the scientific literature,” *Nat. Commun.*, vol. 9, no. 1, p. 2818, Dec. 2018, doi: 10.1038/s41467-018-04976-1.
[115] D. C. Tărlungeanu and G. Novarino, “Genomics in neurodevelopmental disorders: an avenue to personalized medicine,” *Exp. Mol. Med.*, vol. 50, no. 8, Art. no. 8, Aug. 2018, doi: 10.1038/s12276-018-0129-7.
[116] C. M. Dias and C. A. Walsh, “Recent Advances in Understanding the Genetic Architecture of Autism,” *Annu. Rev. Genomics Hum. Genet.*, vol. 21, no. 1, pp. 289–304, 2020, doi: 10.1146/annurev-genom-121219-082309.
[117] N. V. Bryce *et al.*, “Brain parcellation selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity,” *NeuroImage*, vol. 243, p. 118487, Nov. 2021, doi: 10.1016/j.neuroimage.2021.118487.
[118] Y.-M. Kim, J.-B. Poline, and G. Dumas, “Experimenting with reproducibility: a case study of robustness in bioinformatics,” *GigaScience*, vol. 7, no. 7, p. giy077, Jul. 2018, doi: 10.1093/gigascience/giy077.

[119] P. M. Thompson *et al.*, “ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide,” *NeuroImage*, vol. 145, pp. 389–408, 2017, doi: 10.1016/j.neuroimage.2015.11.057.
[120] R. Dimitrova *et al.*, “Phenotyping the preterm brain: characterising individual deviations from normative volumetric development in two large infant cohorts,” Neuroscience, preprint, Aug. 2020. doi: 10.1101/2020.08.05.228700.
[121] R. Dimitrova *et al.*, “Preterm birth alters the development of cortical microstructure and morphology at term-equivalent age,” *NeuroImage*, vol. 243, p. 118488, Nov. 2021, doi: 10.1016/j.neuroimage.2021.118488.
[122] S. Caton and C. Haas, “Fairness in Machine Learning: A Survey,” *ArXiv201004053 Cs Stat*, Oct. 2020, Accessed: Mar. 07, 2022. [Online]. Available: http://arxiv.org/abs/2010.04053
[123] V. Mhasawade, Y. Zhao, and R. Chunara, “Machine learning and algorithmic fairness in public and population health,” *Nat. Mach. Intell.*, vol. 3, no. 8, Art. no. 8, Aug. 2021, doi: 10.1038/s42256-021-00373-4.
[124] E. E. Lee *et al.*, “Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom,” *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 6, no. 9, pp. 856–864, Sep. 2021, doi: 10.1016/j.bpsc.2021.02.001.
[125] B. K. Beaulieu-Jones *et al.*, “Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians?,” *Npj Digit. Med.*, vol. 4, no. 1, pp. 1–6, Mar. 2021, doi: 10.1038/s41746-021-00426-3.
[126] I. Boscolo Galazzo *et al.*, “Explainable Artificial Intelligence for Magnetic Resonance Imaging Aging Brainprints: Grounds and challenges,” *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 99–116, Mar. 2022, doi: 10.1109/MSP.2021.3126573.
[127] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, “The false hope of current approaches to explainable artificial intelligence in health care,” *Lancet Digit. Health*, vol. 3, no. 11, pp. e745–e750, Nov. 2021, doi: 10.1016/S2589-7500(21)00208-9.
[128] E. Klingler, F. Francis, D. Jabaudon, and S. Cappello, “Mapping the molecular and cellular complexity of cortical malformations,” *Science*, vol. 371, no. 6527, Jan. 2021, doi: 10.1126/science.aba4517.
[129] E. Courchesne, T. Pramparo, V. H. Gazestani, M. V. Lombardo, K. Pierce, and N. E. Lewis, “The ASD Living Biology: from cell proliferation to clinical phenotype,” *Mol. Psychiatry*, vol. 24, no. 1, pp. 88–107, Jan. 2019, doi: 10.1038/s41380-018-0056-y.
[130] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nat. Med.*, vol. 25, no. 1, Art. no. 1, Jan. 2019, doi: 10.1038/s41591-018-0300-7.