CHAPTER

1

Introduction to machine learning

Sandra Vieira[1], *Walter Hugo Lopez Pinaya*[1,2], *Andrea Mechelli*[1]

[1] Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom; [2] Centre of Mathematics, Computation, and Cognition, Universidade Federal do ABC, Santo André, São Paulo, Brazil

1.1 Introduction

The field of artificial intelligence is concerned with the theory and implementation of computer systems able to perform tasks that would typically require human intelligence. It relies on the intersection of several disciplines including computer science, engineering, mathematics, statistics, psychology, and neuroscience. Machine learning is an area of artificial intelligence that has emerged as part of the ongoing quest for building intelligent machines that are capable of learning. In this chapter, we introduce machine learning by first highlighting how theories of human learning have inspired its development. When machine learning was first developed, it was heavily informed by the latest theories of human learning from the fields of psychology and neuroscience. This influence has weakened over time, as machine learning gradually grew into a subfield of artificial intelligence after shifting its focus from building cognitive and biological plausible models of learning to solve well-defined learning tasks (Langley, 2011, 2016). However, as we will see in this chapter, several of the fundamental concepts of machine learning are still heavily reliant on theories of human learning. Although the term "machine learning" was coined in 1959 (Samuel, 1959), machine learning only emerged as an area of artificial intelligence in the 1980s (Langley, 2011, 2016). Perhaps as a result of its short history, combined with its interdisciplinary nature, there has been much debate about the definition

https://doi.org/10.1016/B978-0-12-815739-8.00001-8

of machine learning. In this chapter, we present one of the most widely accepted definitions and discuss how machine learning could be applied to brain disorders, both in terms of research and clinical translation. Machine learning is an evolving area that includes a multitude of different approaches. Navigating the literature can easily become overwhelming, especially for the nonexpert. In the final section of this chapter, therefore, we introduce a high-level taxonomy of the main approaches used in the machine learning literature. A good understanding of this taxonomy is a paramount first step toward being able to select the most appropriate machine learning method for a given investigation.

1.2 From human learning to machine learning

The ability to learn is one of the fundamental attributes of intelligent behavior (Langley, 2016). As humans, we are hardwired to learn through our interactions with the environment from the day we are born or even before (James, 2010). It is therefore unsurprising that, when the quest for building intelligent machines began, an important source of inspiration came from what we know about human learning (Michalski, Carbonell, & Mitchell, 1983; Valiant, 1984). Since the beginning of the 20th century, psychology and neuroscience have made significant contributions to our understanding of how humans learn. While a comprehensive account of these contributions is outside the scope of this introductory chapter, below we provide a brief overview of the historical and contextual framework that inspired some of the fundamental ideas that underlie modern machine learning.

Behaviorism was one of the most influential learning theories during the first half of the 20th century. The behaviorist movement spanned over several decades, including Pavlov's initial work on classical conditioning (Pavlov, 1897), Watson's "behaviorist manifesto," which emphasized the importance of studying learning by focusing on observable and measurable behaviors (Watson, 1913), and Skinner's later work on operant conditioning (Skinner, 1953). Central to these approaches and fundamental to machine learning are several concepts including stimulus-response, learning by trial and error, and reinforcement. For example, Pavlov's influential work on stimulus-response was used by Hebb (1949) to show that when an input neuron fires, if it frequently leads to the firing of the output neuron, the synapse between the two is strengthened. This Hebbian rule provided the biological basis for the development of a learning algorithm that became widely used in machine learning models. Likewise, Skinner's work on operational conditioning directly inspired one of the most promising machine learning approaches—reinforcement learning—in which learning emerges from the interaction with the environment,

initially through trial and error and then adjusted through rewards and punishments.

As a reaction to radical behaviorism, more emphasis was given to cognition and their role in learning. Of particular importance to machine learning was the seminal work on category or concept learning by Brunauer and colleagues (Bruner, Goodnow, & Austin, 1956). Briefly, Brunauer and colleagues showed that humans learn to categorize objects by searching and extracting attributes or features to generate an initial rule; this rule is sequentially modified with cumulative exposure to new examples, which do not conform to this rule. This finding led to influential theories about categorization and concept learning in cognitive psychology that have provided the basis for subsequent developments in artificial intelligence (Ashby & Ell, 2001; Ashby & Maddox, 2005, 2011; Weiner et al., 2017). These theories are still reflected in how current machine learning methods perform the categorization of input data into predefined labels or categories. For example, in most neuroimaging studies of brain disorders, exemplar data are used to train a machine learning algorithm to learn the optimal decision boundary that discriminates between brain scans belonging to different groups (e.g., patients and healthy controls). Similarly to human learning, therefore, learning from examples is also one of the cornerstones of current applications of machine learning to brain disorders.

Following the initial influence of behavioral and cognitive theories of learning, a rather different approach inspired by the neurosciences emerged, which involved the use of computational models mimicking the basic characteristics of how neurons receive, transform, and output information (e.g., Fukushima, 1988; Grossberg, 1976; Rosenblatt, 1958). After a long history (Goodfellow, Bengio, Courville, & Bengio, 2016; Schmidhuber, 2015), early attempts based on the functioning of a single neuron evolved into complex models known as artificial neural networks or more recently as deep learning networks, which loosely try to emulate how the brain processes information (LeCun, Bengio, & Hinton, 2015). For example, current state-of-the-art machine learning approaches to visual recognition are based on a specific type of deep learning networks inspired by seminal work on how the primary visual cortex processes information (Hubel & Wisel, 1962). Deep learning has become one of the most popular approaches in machine learning since the early 2000s after record-breaking performances in classical machine learning problems such as object and speech recognition (LeCun et al., 2015). For this reason, this book has dedicated Chapters 9, 10, and 11 to this family of machine learning methods, with each chapter covering a different type of deep learning approach.

In short, the emergence and evolution of machine learning over several decades was meaningfully informed by dominant paradigms from the

fields of psychology and neuroscience. Although perhaps to a lesser extent, this influence is still ongoing, as new computational models of brain function can be used to define new learning algorithms. In an interesting reversal of roles, recent years have mostly seen insights from the growing field of machine learning being integrated within the disciplines of psychology and neuroscience (Bzdok, 2017; Bzdok & Meyer-Lindenberg, 2018; Bzdok and Yeo, 2017).

1.3 What is machine learning?

There is no universal definition of machine learning. Nevertheless, machine learning is usually referred to as an area of artificial intelligence that is concerned with identifying patterns from data and to use these patterns to make predictions about unseen data. In one of the most widely accepted definitions, machine learning is defined as follows:

> "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." ***(Mitchell, 1997, pp. 2).***

In other words, machine learning is "concerned with the question of how to construct computer programs that automatically improve with experience." (Mitchell, 1997, pp. xv). This definition means that, for example, predicting the likelihood that individuals with mild cognitive impairment (MCI) will develop Alzheimer's disease (AD) later in life (T) can be achieved by using relevant input data (e.g., demographic, neuroimaging, genetic, neurocognitive, and clinical data) from several individuals (E); if the algorithm has successfully learned (P), it will then be capable of using these data to predict clinical outcome of new MCI individuals.

The main outcome of machine learning is a measure (proxy) of generalizability: the extent to which the model is capable of outputting correct predictions when presented with new data, based on learned rules from previous exposure to similar (but not the same) data (Domingos, 2012). Regardless of how much data we can expose the algorithm to, it is most unlikely that it will see the same data again; for this reason, it is crucial that the model learns patterns that are generalizable to make accurate predictions in new data. For example, in our earlier example involving people with MCI who do and do not develop AD, the algorithm should learn differences between the two groups that are intrinsic to clinical outcome regardless of the characteristics of the cohort under investigation. In contrast, if the algorithm is based on spurious differences that are specific to one cohort, it will not be possible to apply it to the rest of the MCI population. As we will repeatedly see throughout this book,

achieving high levels of generalizability is a key challenge in machine learning.

1.4 How is machine learning relevant to brain disorders?

1.4.1 From group-level to individual-level inferences

The last three decades have seen impressive advances in the understanding of the neurobiological basis of brain disorders, including both psychiatric and neurological disease. However, very few results have been incorporated into clinical practice (Dazzan, 2014; Prata, Mechelli, & Kapur, 2014; Woo, Chang, Lindquist, & Wager, 2017). One of the main reasons for this gap between research and clinical practice is that the former has been dominated by methods that only allow inferences at group level (e.g., a group of psychosis patients have larger ventricles than a group of controls), while a clinician has to make diagnostic or treatment decisions at the level of the individual. A key reason why machine learning is gaining considerable attention among the research and medical communities is that it promises to bridge this gap. By learning patterns in the data that best distinguish between patients with a certain disease of interest and healthy people, for example, it is possible to estimate the likelihood that a new set of data acquired from an individual belongs to a patient or a healthy person. Similarly, by learning patterns in the data that best distinguish between patients who benefit from a specific treatment and patients who do not benefit from it, it is possible to estimate the likelihood that a new set of data acquired from an individual belongs to a "responder" or a "nonresponder." Therefore, machine learning opens new possibilities in personalized medicine, by allowing the development of novel tools that could be used to inform diagnostic and treatment decision-making in everyday clinical practice.

1.4.2 From univariate to multivariate inferences

The vast majority of neuroscientific studies of brain disorders are based on mass-univariate methods, i.e., a separate statistical test is performed to investigate each variable of interest. In neuroimaging, for example, statistical parametric mapping is used to perform a large number of voxel-wise comparisons between groups, without considering possible interaction between voxels. However, this approach is not in line with the current understanding of brain function in health and disease (Biswal et al., 2010; Fox et al., 2005); for example, several psychiatric and neurological symptoms are best explained by subtle and widespread network-level changes in structure and function rather than focal alternations

(Kennedy & Courchesne, 2008; Mulders, van Eijndhoven, Schene, Beckmann, & Tendolkar, 2015; Nielsen et al., 2013; Sheffield & Barch, 2016). Although traditional statistics do offer multivariate solutions (e.g., multivariate analysis of variance [MANOVA]), these have not been widely used in the neuroscientific literature. A further reason why machine learning is gaining considerable attention, therefore, is that it is inherently a multivariate approach. In other words, it is capable of taking the relationship between multiple variables inputted into the same model into account, thereby allowing greater sensitivity to subtle and widespread network-level changes in brain structure and function. A further advantage of a multivariate approach is that it allows one to combine different types of data (e.g., neuropsychological tests and neuroimaging) within the same statistical model, which can enhance prediction (e.g., Moradi, Pepe, Gaser, Huttunen, & Tohka, 2015; Pettersson-Yeo et al., 2014; Wu et al., 2017).

1.4.3 Focus on prediction and generalizability

Research into brain disorders has historically relied on classical inferential statistics, which is mainly concerned with elucidating the relationship between observed phenomena of interest (Yarkoni & Westfall, 2017). In neuroimaging, for example, studies typically use the general linear model, which estimates the strength of the association between independent and dependent variables to generate a measure of explained variance or goodness of fit (i.e., the extent to which the variance in the independent variable(s) explains the variance in the dependent variable(s)) (Friston et al., 1994). It is often assumed that models with high explanatory power or goodness of fit have high predictive power. However, from a statistical perspective, the model that best describes a set of observations will not necessarily be the most successful at predicting real-world outcomes (Arbabshirani, Plis, Sui, & Calhoun, 2017; Shmueli, 2010). A detailed explanation of the discrepancy between classical statistics and predictive power is beyond the scope of this chapter (see Bzdok (2017) and Yarkoni & Westfall (2017) for a comprehensive explanation). One of the main reasons for this discrepancy is that, while a statistical model may achieve a high explanatory power or goodness of fit when fitted to a particular dataset, it will likely incorporate the unique characteristics of the dataset (Yarkoni & Westfall, 2017). When a dataset is large enough to ensure representativeness of the population from which it was drawn, this is less likely to be an issue. However, in the vast majority of brain disorders research, which involves studies with small samples, representativeness is not guaranteed, and models will likely capture fluctuations in the data that are unique to a particular sample; this is known as "overfitting," a critical

concept in machine learning that will be discussed in greater detail in Chapter 2. The extent to which findings from a single study are generalizable and used to make predictions in other samples is not usually addressed in studies using classical statistics (Bzdok & Yeo, 2017). These studies tend to use all available data in a single statistical analysis, for example, by taking all observations from the healthy control group and comparing them to all observations from the patient group, without testing their models on an independent set of data (Bzdok, 2017).

Importantly, machine learning does not necessarily solve the issue of generalizability. However, it does at least attempt to measure it. Indeed, building models capable of accurate predictions in unseen data is the fundamental goal of machine learning. It is this shift from maximizing explanatory power toward maximizing generalizability that makes machine learning fundamentally different from traditional research based on classical statistics. As we will see throughout this book, building generalizable models is a critical challenge in brain disorders research where difficult questions are often addressed with complex data (e.g., high-dimensional data, multimodal data) and small samples. However, focusing on prediction and generalizability is indispensable in the long-term goal of bringing personalized medicine to brain disorders.

1.4.4 Data-driven models and big data

Historically, research into brain disorders has been heavily based on deductive (top-down) or theory-driven approaches, where carefully thought-out and well-defined hypotheses are tested and ultimately confirmed or rejected. Having a priori hypothesis is considered paramount as it minimizes the risk of false-positive findings and *post hoc* explanations (Kitchin, 2014). More recently, increasing access to large datasets combined with technological advances has propelled the emergence of data-driven approaches, such as machine learning, where insights are generated purely from data in a bottom-up fashion. Contrary to classical statistics, where the aim is to test a priori hypotheses while making significant assumptions about the data, in machine learning, the main premise is to "let the data speak for themselves," while making as few assumptions as possible about the data (Bzdok, 2017; Jordan & Mitchell, 2015; Mahmoodi, Leckelt, van Zalk, Geukes, & Back, 2017). For example, some of the state-of-the-art models for facial recognition were developed by feeding millions of images to an algorithm that learns to recognize faces without being given any a priori information about what a face looks like (Taigman, Yang, Ranzato, & Wolf, 2014). In theory, having millions of observations is not a requirement for machine learning. However, it is with large amounts of data that machine learning thrives, particularly the most cutting-edge approaches (LeCun et al., 2015).

Although big data is not yet common in brain disorders, the research community is becoming increasingly aware of its advantages, leading to a growing number of concerted efforts (Iniesta, Stahl, & Mcguffin, 2016; Mahmoodi et al., 2017; Poldrack & Gorgolewski, 2014). For instance, the development of neuroimaging and genetic consortiums, such as ENIGMA (Bearden & Thompson, 2017), ADNI (Mueller et al., 2005), UK Biobank (Sudlow et al., 2015), and the Psychiatric Genomics Consortium (Lee et al., 2013), has resulted in unprecedented sample sizes (in the order of thousands) in schizophrenia (van Erp et al., 2018), bipolar disorder (Hibar et al., 2018), major depressive disorder (Schmaal et al., 2017; Shen et al., 2017), AD (Weiner et al., 2017), and autism (Weiner et al., 2017). Data sharing initiatives are also increasing rapidly, with over 40 online repositories for neuroscientific data in 2015 (Eickhoff, Nichols, Van Horn, & Turner, 2016; Ferguson, Nielson, Cragin, Bandrowski, & Martone, 2014). Moving forward, as the research community continues to invest in the pursuit of large datasets, machine learning is likely to keep gaining momentum (Bzdok & Yeo, 2017). Initial attempts to apply machine learning to larger samples have been performed in bipolar disorder using data from 3020 participants (Nunes et al., 2018), in schizophrenia with 941 individuals (Rozycki et al., 2018), and in AD with 20,000 participants (Wegmayr, Aitharaju, & Buhmann, 2018).

1.4.5 Recognizing heterogeneity

It is well known that psychiatric and neurological disorders tend to be heterogeneous concerning underlying neuroanatomical and neurofunctional alterations, clinical presentation, and progression over time (Holmes & Patrick, 2018; Insel et al., 2010; Wardenaar & de Jonge, 2013). However, most advances in brain disorders research, including rigorous clinical trials, for example, are based on the idea of a "typical patient," which ignores individual variability (Wolfers et al., 2018). In contrast, machine learning involves the examination of individual differences in the data and the use of this information to make inferences at the level of the individual. However, it has to be acknowledged that integrating heterogeneity in individual-level modeling can be challenging, as it becomes more difficult to find patterns in the data that are relevant to the task above and beyond individual heterogeneity (Schnack, 2017). In light of the current trend to recruit larger and larger sample sizes, the data are likely to become more heterogeneous, in contrast with the traditional case−control approach where, ideally, both the patient group and the control group are expected to be as homogeneous as possible. On the other hand, larger samples are likely to be more representative of the illness and thus carry more translational potential in clinical practice. The issue of heterogeneity in machine learning is gaining increasing attention in brain disorders research, and it will be explored in detail in Chapter 14.

1.5 Different types of machine learning

As we saw in the previous sections, machine learning involves learning relevant patterns in the data and then using them to make predictions. There are multiple ways in which this can be achieved, resulting in a multitude of machine learning algorithms to choose from. A common taxonomy organizes the different approaches according to the style of learning. Based on this categorization, machine learning methods are grouped into four different types of learning: supervised, unsupervised, semisupervised, and reinforcement learning (Fig. 1.1). Supervised learning is by far the most commonly used approach in general and in brain disorders research and therefore will be discussed in greater detail.

1.5.1 Supervised learning

In supervised learning, the algorithm has access to what it is trying to predict, i.e., the target variable; this could be, for example, the presence or absence of illness, severity of symptoms, or future clinical outcome. The aim here is to use an algorithm to learn the optimal function that best captures the relationship between the input and the target variable. The reason this type of learning is called "supervised" is that the algorithm has prior knowledge of what the output values should be (e.g., patient with depression vs. healthy control). The algorithm is trained using multiple examples and is allowed to receive feedback during the learning

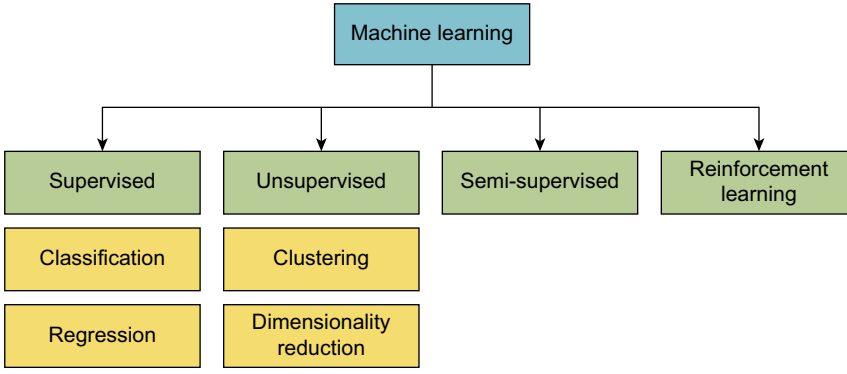

FIGURE 1.1 Types of learning in machine learning. Learning can be supervised, unsupervised, semisupervised, or alternatively through reinforcement learning. Supervised learning can be further grouped into classification or regression task depending on whether the true target is categorical or continuous. The main applications of unsupervised learning in brain disorders include clustering and reducing the dimensionality of the input data.

process based on how close its prediction matches the true target. This type of learning is often compared to learning with a teacher: the teacher knows the correct answers and corrects the algorithm when it makes a mistake. In this context, therefore, learning is an iterative process of predictions and subsequent adjustments, until the difference between output predictions and target is minimized as much as possible. Performance is measured by comparing the algorithm's predictions against the true target values in unseen data. Depending on whether the target variable is a categorical or continuous variable, a supervised learning task is either classification or regression problem, respectively.

1.5.1.1 Classification

Classification algorithms aim to predict group membership, also known as labels or classes, for a set of observations. It is the most commonly used type of algorithm in brain disorders research. The popularity of this type of algorithm can be explained by the fact that most clinical problems can be simplified into a categorical decision; for instance, should a certain patient be treated with medication A, B, or C? The most straightforward application of a classifier in brain disorders is diagnostic classification. In this type of task, an algorithm learns to distinguish, for example, patients with a particular disease from healthy controls. The overwhelming majority of diagnostic classification studies published so far have used neuroimaging data (for reviews on different disorders, see Kambeitz et al., 2017, 2015; Kim & Na, 2018; Rathore, Habes, Iftikhar, Shacklett, & Davatzikos, 2017; Retico, Tosetti, Muratori, & Calderoni, 2014; Woo et al., 2017). A smaller number of studies have used other types of data, including, for example, motor symptoms to identify individuals suffering from Parkinson's disease (Ahlrichs & Lawo (2013) for a review) and genetic information to identify schizophrenia patients (Aguiar-Pulido et al., 2010; Yang, Liu, Sui, Pearlson, & Calhoun, 2010). This approach can be easily extended to differential diagnosis, where an algorithm can predict the probability of a particular patient belonging to each of several diagnostic categories (e.g., Koutsouleris et al., 2015; Schnack et al., 2014). Furthermore, classification can be adapted to distinguish between subtypes within a specific disorder (e.g., Qureshi, Min, Jo, & Lee, 2016). A further application of a classifier in brain disorders is prognostic classification, which involves the prediction of longitudinal outcomes such as illness transition, illness progression, or treatment response in a particular patient. As an example, a few studies have now attempted to predict conversion from MCI to AD (Pellegrini et al., 2018) and from prodromal to full psychosis (Janssen, Mourão-Miranda, & Schnack, 2018; McGuire et al., 2015) using neuroimaging and neuropsychological data.

1.5.1.2 Regression

In a regression problem, the aim is to predict a score on a continuous scale. Several questions that can be addressed with a classifier can also be answered using a regression algorithm (which is also known as regressor in the literature), by defining the outcome as a continuous rather than a categorical variable. This characteristic can be useful when, for example, one is predicting outcomes such as treatment response or illness progression which fall along a continuum.

For instance, Mechelli et al. (2017) used both a classifier and a regressor to predict the level of functioning in individuals at high risk for psychosis from clinical data acquired 2 years earlier. Here, level of functioning was first measured on a continuous score. For the classification task, functioning was operationalized using a threshold for distinguishing between poor and good functioning from the continuous scores and data were analyzed with support vector machine, a common classification algorithm; for the regression task, functioning was operationalized using the continuous score and analyzed with support vector regression, a common regression algorithm. Other examples of regressors include using neuroanatomical information to predict clinical scores in individuals suffering from AD (Stonnington et al., 2010; Wang, Fan, Bhatt, & Davatzikos, 2010) or autism spectrum disorders (Sato et al., 2013) or using brain activation to predict symptoms progression in Huntington's disease (Rizk-Jackson et al., 2011).

1.5.2 Unsupervised learning

Unlike supervised learning, in an unsupervised learning task, there is no target value. The aim is rather to uncover underlying structures in the data. There are two main applications of unsupervised learning useful to brain disorders research: clustering and dimensionality reduction.

1.5.2.1 Clustering

Cluster analysis is an analytical technique for developing meaningful subgroups from a larger sample, such as subgroups of patients with different clinical profiles within a larger patient cohort. Here, individuals are classified into a smaller number of mutually exclusive and not predefined groups based on the observed similarities among them (Hair, Black, Babin, & Anderson, 2014). Exemplar applications of clustering in brain disorders research include investigating the neurocognitive profile of bipolar disorders patients (Wu et al., 2017) or identifying patients with and without autism spectrum disorders from electronic health records (Lingren et al., 2016). A more detailed discussion of this analytical technique can be found in Chapter 13.

1.5.2.2 *Dimensionality reduction*

Dimensionality reduction is useful in situations when the number of features is substantially larger than the number of observations. This is a common problem in neuroimaging, where typically less than 100 participants are recruited, and each brain scan contains several thousands of voxels. As a result, the number of features (voxels) greatly outnumbers the number of observations (sample size). This problem is known as the curse of dimensionality (Bellman, 1961) or "small n, large p" (Fort & Lambert-Lacroix, 2005). In such instances, reducing the number of features can be beneficial to alleviate computational demands, remove redundant or irrelevant information, and reduce the risk of overfitting (Guyon & Elisseeff, 2003). Several techniques can be used to perform dimensionality reduction, such as principal component analysis (PCA) (Lever, Krzywinski, & Altman, 2017), independent component analysis (McKeown & Sejnowski, 1998), or autoencoders (Vincent, Larochelle, Lajoie, Bengio, & Manzagol, 2010). PCA is arguably one of the most common approaches to dimensionality reduction in brain disorders and will be discussed in detail in Chapter 12. Autoencoders have not been commonly used in the past; however, they are becoming increasingly popular and will be covered in more detail in Chapter 11.

1.5.3 Semisupervised learning

As the name suggests, in semisupervised learning, the labels or target variables are only available for a portion of the data (Zhu & Goldberg, 2009). Semisupervised learning addresses this problem by allowing the model to integrate the available unlabeled data in its supervised learning. This approach is useful when it is impractical or too expensive to access or measure the target variable for all participants. For example, longitudinal studies investigating illness progression that require several years of follow-up to obtain a reliable disease label (e.g., continuous, episodic or intermediate psychotic illness course) can be impractical or prohibitively costly. In such cases, semisupervised learning can be used to model the existing labeled data, i.e., patients who have completed their participation in the study, as well as unlabeled data, i.e., patients for whom a reliable label has not yet been established. This approach maximizes the amount of input data in the context of studies with finite resources and time. While the use of semisupervised learning is not common in brain disorders, there have been a few studies using this approach. These studies have reported better results when both labeled and unlabeled data were utilized within a semisupervised framework compared to the use of labeled data only within a supervised context (Moradi et al., 2015).

1.5.4 Reinforcement learning

In reinforcement learning, the aim is to build a system that can learn from interacting with the environment, much like in operant conditioning (Sutton & Barto, 1998). In this type of learning, the algorithm's behavior is shaped through a sequence of rewards and penalties, which depend on whether its decisions toward a defined goal are correct or incorrect, as defined by the researcher. Unlike supervised learning, where the algorithm uses the examples given to model behavior, in reinforcement learning the algorithm is allowed to behave freely, i.e., on the basis of trial and error, to discover what actions maximize reward and minimize penalty. Reinforcement learning is one of the most promising areas of machine learning across a range of disciplines. However, applications of this type of learning in brain disorders have been very limited so far, and therefore a detailed explanation of this type of learning is beyond the scope of this book.

1.6 Conclusion

Machine learning is an emerging area of artificial intelligence that is gaining momentum within brain disorders research. While its rapid expansion and multidisciplinary nature make it difficult to extract a single and comprehensive definition from the existing literature, it is agreed that machine learning is mainly concerned with identifying patterns from data and using these patterns to make predictions about unseen data. This definition is in sharp contrast with classical inferential statistics, where the aim is to maximize explanatory power. Also, contrary to classical statistics, machine learning is capable of making inferences at the individual level. It is this shift from the "typical" to the "individual" patient that makes machine learning so promising in the brain sciences. It is also critical to recognize that there are important challenges in the application of machine learning to brain disorders. For example, as a data-driven approach, machine learning requires large numbers of observations, whereas brain disorders research traditionally relies on small sample sizes. This challenge is now being addressed through increasing numbers of sharing initiatives and multicenter consortia, which have already led to unprecedented sample sizes in several brain disorders.

Machine learning is a growing area with a multitude of methods to choose from. In this chapter, we have introduced a high-level categorization of these methods based on learning style. As supervised learning is overwhelmingly the most common type of machine learning in brain disorders research, the next two chapters provide a deeper understanding

of this type of approach. Chapter 2 provides an overview of a standard supervised learning pipeline and introduces some of the fundamental concepts in machine learning that will be used throughout the book. Chapter 3 illustrates some of the most recent advances in the supervised machine learning literature by discussing in detail a few studies aiming at answering prominent questions in brain disorders.

1.7 Key points

- Machine learning is an area of artificial intelligence that has emerged as part of the ongoing quest for building intelligent machines capable of learning.
- Prominent theories of human learning from psychology and neuroscience, such as behaviorism, cognitivism, and neurophysiology, greatly influenced the early stages of machine learning.
- Machine learning differs from traditional statistics in at least four ways: it is capable of inferences at the individual level; it focuses on maximizing generalizability; it takes into account individual heterogeneity; and it is a data-driven approach.
- There are four main types of machine learning according to the style of learning: supervised, unsupervised, semisupervised, and reinforcement learning.
- Supervised learning is the most common type of machine learning in brain disorders research.

References

Aguiar-Pulido, V., Seoane, J. A., Rabuñal, J. R., Dorado, J., Pazos, A., Munteanu, C. R., et al. (2010). Machine learning techniques for single nucleotide polymorphism—disease classification models in schizophrenia. *Molecules, 15*(7), 4875–4889. https://doi.org/10.3390/molecules15074875.

Ahlrichs, C., & Lawo, M. (2013). Parkinson's disease motor symptoms in machine learning: a review. *Health Informatics-An International Journal (HIIJ), 2*(4). https://doi.org/10.5121/hiij.2013.2401.

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *Neuroimage, 145*, 137–165. https://doi.org/10.1016/j.neuroimage.2016.02.079.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences, 5*(5), 204–210. https://doi.org/10.1016/S1364-6613(00)01624-7.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224*(1), 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x.

Bearden, C. E., & Thompson, P. M. (2017). Emerging global initiatives in neurogenetics: the enhancing neuroimaging genetics through meta-analysis (ENIGMA) consortium. *Neuron, 94*(2), 232–236. https://doi.org/10.1016/j.neuron.2017.03.033.

Bellman, R. (1961). *"Curse of dimensionality." Adaptive Control Processes: A Guided Tour*. Princeton, NJ.

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America, 107*(10), 4734–4739. https://doi.org/10.1073/pnas.0911855107.

Bruner, J. S., Goodnow, J. J., & Austin, G. (1956). *A Study of Thinking*. New York.

Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Frontiers in Neuroscience, 11*, 543. https://doi.org/10.3389/fnins.2017.00543.

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(3), 223–230. https://doi.org/10.1016/J.BPSC.2017.11.007.

Bzdok, D., & Yeo, B. T. T. (2017). Inference in the age of big data: future perspectives on neuroscience. *Neuroimage, 155*, 549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061.

Dazzan, P. (2014). Neuroimaging biomarkers to predict treatment response in schizophrenia: the end of 30 years of solitude? *Dialogues in Clinical Neuroscience, 16*(4), 491–503.

Domingos, P. (2012). A few useful things to know about machine learning. *Commun. Acm, 55*(10), 78–87. https://doi.org/10.1145/2347736.2347755.

Eickhoff, S., Nichols, T. E., Van Horn, J. D., & Turner, J. A. (2016). Sharing the wealth: neuroimaging data repositories. *Neuroimage, 124*(Pt B), 1065–1068. https://doi.org/10.1016/j.neuroimage.2015.10.079.

Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., & Martone, M. E. (2014). Big data from small data: data-sharing in the "long tail" of neuroscience. *Nature Publishing Group, 17*(11), 1442. https://doi.org/10.1038/nn.3838.

Fort, G., & Lambert-Lacroix, S. (2005). Classification using partial least squares with penalized logistic regression. *Bioinformatics, 21*(7), 1104–1111. https://doi.org/10.1093/bioinformatics/bti114.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673–9678.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping, 2*(4), 189–210. https://doi.org/10.1002/hbm.460020402.

Fukushima, K. (1988). Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Networks, 1*(2), 119–130.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep Learning* (vol. 1). Cambridge: MIT press.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23*(3), 121–134.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(Mar), 1157–1182.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis*. Harlow : Pearson Education Limited.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Hibar, D. P., Westlye, L. T., Doan, N. T., Jahanshad, N., Cheung, J. W., Ching, C. R. K., et al. (2018). Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular Psychiatry, 23*(4), 932–942. https://doi.org/10.1038/mp.2017.73.

Holmes, A. J., & Patrick, L. M. (2018). The myth of optimality in clinical neuroscience. *Trends in Cognitive Sciences, 22*(3), 241–257. https://doi.org/10.1016/j.tics.2017.12.006.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology, 160*(1), 106–154.

Iniesta, R., Stahl, D., & Mcguffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological medicine, 46*(12), 2455–2465. https://doi.org/10.1017/S0033291716001367.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379.

James, D. K. (2010). Fetal learning: a critical review. *Infant and Child Development: An International Journal of Research and Practice, 19*(1), 45–54.

Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(9), 798–808. https://doi.org/10.1016/J.BPSC.2018.04.004.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. *Science (New York, N.Y.), 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415.

Kambeitz, J., Cabral, C., Sacchet, M. D., Gotlib, I. H., Zahn, R., Serpa, M. H., et al. (2017). Detecting neuroimaging biomarkers for depression: a meta-analysis of multivariate pattern recognition studies. *Biological Psychiatry, 82*(5), 330–338. https://doi.org/10.1016/j.biopsych.2016.10.028.

Kambeitz, J., Kambeitz-Ilankovic, L., Leucht, S., Wood, S., Davatzikos, C., Malchow, B., et al. (2015). Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology, 40*(7), 1742–1751. https://doi.org/10.1038/npp.2015.22.

Kennedy, D. P., & Courchesne, E. (2008). The intrinsic functional organization of the brain is altered in autism. *Neuroimage, 39*(4), 1877–1885. https://doi.org/10.1016/j.neuroimage.2007.10.052.

Kim, Y.-K., & Na, K.-S. (2018). Application of machine learning classification for structural brain MRI in mood disorders: critical review from a clinical perspective. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 80*, 71–80.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 205395171452848 https://doi.org/10.1177/2053951714528481.

Koutsouleris, N., Meisenzahl, E. M., Borgwardt, S., Riecher-Rössler, A., Frodl, T., Kambeitz, J., et al. (2015). Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain, 138*(7), 2059–2073. https://doi.org/10.1093/brain/awv111.

Langley, P. (2011). The changing science of machine learning. *Mach Learn, 82*, 275–279. https://doi.org/10.1007/s10994-011-5242-y.

Langley, P. (2016). The central role of cognition in learning. *Advances in Cognitive Systems* (Vol. 4). Retrieved from http://www.cogsys.org/papers/ACSvol4/paper2.pdf.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539.

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics, 45*(9), 984–994. https://doi.org/10.1038/ng.2711.

Lever, J., Krzywinski, M., & Altman, N. (2017). Points of significance: principal component analysis. *Nature Methods, 14*(7), 641–642. https://doi.org/10.1038/nmeth.4346.

Lingren, T., Chen, P., Bochenek, J., Doshi-Velez, F., Manning-Courtney, P., Bickel, J., et al. (2016). Electronic health record based algorithm to identify patients with autism spectrum disorder. *PloS One, 11*(7), e0159621.

Mahmoodi, J., Leckelt, M., van Zalk, M., Geukes, K., & Back, M. (2017). Big Data approaches in social and behavioral science: four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences, 18*, 57–62. https://doi.org/10.1016/j.cobeha.2017.07.001.

McGuire, P., Sato, J. R., Mechelli, A., Jackowski, A., Bressan, R. A., & Zugman, A. (2015). Can neuroimaging be used to predict the onset of psychosis? *The Lancet Psychiatry, 2*(12), 1117–1122.

McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Human Brain Mapping, 6*(5–6), 368–372. https://doi.org/10.1002/(SICI)1097-0193(1998)6:5/6<368::AID-HBM7>3.0.CO;2-E.

Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., et al. (2017). Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. *Schizophrenia Research, 184*, 32–38. https://doi.org/10.1016/j.schres.2016.11.047.

Michalski, R., Carbonell, J., & Mitchell, T. (1983). *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media.

Mitchell, T. M. (1997). Machine learning. 1997. Burr ridge, IL. *McGraw Hill, 45*(37), 870–877.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., & Tohka, J. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage, 104*, 398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). *Alzheimer's & Dementia, 1*(1), 55–66. https://doi.org/10.1016/j.jalz.2005.06.003.

Mulders, P. C., van Eijndhoven, P. F., Schene, A. H., Beckmann, C. F., & Tendolkar, I. (2015). Resting-state functional connectivity in major depressive disorder: a review. *Neuroscience and Biobehavioral Reviews, 56*, 330–344. https://doi.org/10.1016/j.neubiorev.2015.07.014.

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., et al. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience, 7*, 599. https://doi.org/10.3389/fnhum.2013.00599.

Nunes, A., Schnack, H. G., Ching, C. R. K., Agartz, I., Akudjedu, T. N., Alda, M., et al. (2018). Using structural MRI to identify bipolar disorders – 13 site machine learning study in 3020 individuals from the ENIGMA Bipolar Disorders Working Group. *Molecular Psychiatry, 1*. https://doi.org/10.1038/s41380-018-0228-9.

Pavlov, I. P. (1897). *The Work of the Digestive Glands*. London: Griffin.

Pellegrini, E., Ballerini, L., Hernandez, M. del C. V., Chappell, F. M., González-Castro, V., Anblagan, D., et al. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimer's & Dementia: diagnosis, Assessment & Disease Monitoring*. https://doi.org/10.1016/j.dadm.2018.07.004.

Pettersson-Yeo, W., Benetti, S., Marquand, A. F., Joules, R., Catani, M., Williams, S. C. R., et al. (2014). An empirical comparison of different approaches for combining multimodal neuroimaging data with support vector machine. *Frontiers in Neuroscience, 8*, 189.

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nature Neuroscience, 17*(11), 1510–1517. https://doi.org/10.1038/nn.3818.

Prata, D., Mechelli, A., & Kapur, S. (2014). Clinically meaningful biomarkers for psychosis: a systematic and quantitative review. *Neuroscience & Biobehavioral Reviews, 45*, 134–141.

Qureshi, M. N. I., Min, B., Jo, H. J., & Lee, B. (2016). Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study. *PLOS ONE, 11*(8), e0160697. https://doi.org/10.1371/journal.pone.0160697.

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., & Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage, 155*, 530–548. https://doi.org/10.1016/j.neuroimage.2017.03.057.

Retico, A., Tosetti, M., Muratori, F., & Calderoni, S. (2014). Neuroimaging-based methods for autism identification: a possible translational application? *Functional Neurology, 29*(4), 231–239.

Rizk-Jackson, A., Stoffers, D., Sheldon, S., Kuperman, J., Dale, A., Goldstein, J., et al. (2011). Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques. *Neuroimage, 56*(2), 788–796. https://doi.org/10.1016/j.neuroimage.2010.04.273.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386.

Rozycki, M., Satterthwaite, T. D., Koutsouleris, N., Erus, G., Doshi, J., Wolf, D. H., et al. (2018). Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable Across diverse patient populations and within individuals. *Schizophrenia Bulletin, 44*(5), 1035–1044. https://doi.org/10.1093/schbul/sbx137.

Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229.

Sato, J. R., Hoexter, M. Q., Oliveira, P. P. de M., Brammer, M. J., Murphy, D., & Ecker, C. (2013). Inter-regional cortical thickness correlations are associated with autistic symptoms: a machine-learning approach. *Journal of Psychiatric Research, 47*(4), 453–459. https://doi.org/10.1016/J.JPSYCHIRES.2012.11.017.

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N., et al. (2017). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Molecular Psychiatry, 22*(6), 900–909. https://doi.org/10.1038/mp.2016.60.

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Schnack, H. G. (2017). Improving individual predictions: machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2017.10.023.

Schnack, H. G., Nieuwenhuis, M., van Haren, N. E. M., Abramovic, L., Scheewe, T. W., Brouwer, R. M., et al. (2014). Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage, 84*, 299–306. https://doi.org/10.1016/j.neuroimage.2013.08.053.

Sheffield, J. M., & Barch, D. M. (2016). Cognition and resting-state functional connectivity in schizophrenia. *Neuroscience and Biobehavioral Reviews, 61*, 108–120. https://doi.org/10.1016/j.neubiorev.2015.12.007.

Shen, X., Reus, L. M., Cox, S. R., Adams, M. J., Liewald, D. C., Bastin, M. E., et al. (2017). Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank imaging data. *Scientific Reports, 7*(1), 5547. https://doi.org/10.1038/s41598-017-05507-6.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330.

Skinner, B. F. (1953). *Science and Human Behavior*. New York: MacMillan.

Stonnington, C. M., Chu, C., Klöppel, S., Jack, C. R., Ashburner, J., & Frackowiak, R. S. J. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage, 51*(4), 1405–1413. https://doi.org/10.1016/j.neuroimage.2010.03.051.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine, 12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to Reinforcement Learning* (vol. 135). Cambridge: MIT press.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1701–1708).

Valiant, G. (1984). A theory of the learnable. In *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing* (pp. 436–445). ACM.

van Erp, T. G. M., Walton, E., Hibar, D. P., Schmaal, L., Jiang, W., Glahn, D. C., et al. (2018). Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the enhancing neuro imaging genetics through meta analysis (ENIGMA) consortium. *Biological Psychiatry, 84*(9), 644–654. https://doi.org/10.1016/j.biopych.2018.04.023.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research, 11*(Dec), 3371–3408.

Wang, Y., Fan, Y., Bhatt, P., & Davatzikos, C. (2010). High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. *Neuroimage, 50*(4), 1519–1535. https://doi.org/10.1016/j.neuroimage.2009.12.092.

Wardenaar, K. J., & de Jonge, P. (2013). Diagnostic heterogeneity in psychiatry: towards an empirical solution. *BMC Medicine, 11*(1), 201. https://doi.org/10.1186/1741-7015-11-201.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*(2), 158–177. https://doi.org/10.1037/h0074428.

Wegmayr, V., Aitharaju, S., & Buhmann, J. (2018). Classification of brain MRI with big data and deep 3D convolutional neural networks. In K. Mori, & N. Petrick (Eds.), *Medical Imaging 2018: Computer-Aided Diagnosis* (p. 63). SPIE. https://doi.org/10.1117/12.2293719.

Weiner, D. J., Wigdor, E. M., Ripke, S., Walters, R. K., Kosmicki, J. A., Grove, J., et al. (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature Genetics, 49*(7), 978–985. https://doi.org/10.1038/ng.3863.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2017). Recent publications from the Alzheimer's Disease Neuroimaging Initiative: reviewing progress toward improved AD clinical trials. *Alzheimer's & Dementia, 13*(4), e1–e85. https://doi.org/10.1016/j.jalz.2016.11.007.

Wolfers, T., Doan, N. T., Kaufmann, T., Alnæs, D., Moberget, T., Agartz, I., et al. (2018). Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA psychiatry, 75*(11), 1146–1155. https://doi.org/10.1001/jamapsychiatry.2018.2467.

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience, 20*(3), 365–377. https://doi.org/10.1038/nn.4478.

Wu, M.-J., Mwangi, B., Bauer, I. E., Passos, I. C., Sanches, M., Zunta-Soares, G. B., et al. (2017). Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. *Neuroimage, 145*, 254–264. https://doi.org/10.1016/j.neuroimage.2016.02.016.

Yang, H., Liu, J., Sui, J., Pearlson, G., & Calhoun, V. D. (2010). A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. *Frontiers in Human Neuroscience, 4*, 192. https://doi.org/10.3389/fnhum.2010.00192.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 3*(1), 1–130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006.