Preface

Across several areas of medicine, the traditional one-size-fits-all approach to diagnosis and treatment is being replaced with a more nuanced approach that recognizes that we are all individuals. This paradigm shift is also taking place in psychiatry and neurology, where there is increasing recognition that each individual carries a unique combination of strengths and vulnerabilities. People differ not only with respect to the risk of developing an illness in the first place but also with respect to the chance of making a good recovery over time, depending on biological, psychological, and social factors. Therefore, to develop disease models that can be used to optimize diagnosis and treatment, we need statistical techniques that allow us to capture individual differences along these factors. This is where machine learning comes into play. While traditional analytical methods based on classical statistics allow inferences at group level, and as such have limited clinical applicability, machine learning methods allow inferences at the level of the individual and therefore enable a more personalized approach to diagnosis and treatment. In light of this critical advantage, together with the ability to capture multivariate relationships in the data, machine learning methods are becoming increasingly popular in the investigation of brain disorders. A cursory search for "machine learning" in "psychiatric" or "neurological" disorders in the scientific literature brings up 16 articles in 2008, 62 articles in 2013, and 254 articles in 2018—an exponential increase that illustrates the growing interest in these methods. Perhaps for the first time, researchers and clinicians with a shared interest in brain disorders are working together to develop and validate tools than could be used to assist the diagnosis and treatment of individual patients. This endeavor, however, will only succeed if both researchers and clinicians are aware of the nature of the machine learning approach, including its methodological challenges and potential advantages and disadvantages. For example, good appreciation of the machine learning pipeline is required for interpreting the results and avoiding spurious conclusions. Yet, currently available textbooks on state-of-the art machine learning methods tend to be written in a highly technical language and can be challenging to understand for researchers and clinicians without an academic background in engineering, mathematics, or statistics.

This book is aimed at researchers and clinicians interested in the application of machine learning to psychiatry and neurology—two

related medical areas that here we combine under the umbrella term of "brain disorders." The book is written for a nontechnical audience, such as psychologists, neuroscientists, psychiatrists, neurologists, and other health-care practitioners, and assumes no prior knowledge of machine learning (although some understanding of statistics is desirable). It should be acknowledged that this book does not cover all available machine learning methods; instead, we focus on the most widely used techniques, as well as the most novel and promising ones, in the field of brain disorders specifically. Our aim is to provide the reader with the knowledge and understanding required to select and implement machine learning models that can be used to address the main research and clinical challenges in brain disorders. These fall within three broad categories: (i) prediction of illness—studies attempting to predict whether people at clinical risk for a certain disease will or will not become ill in the future; (ii) diagnostic evaluation—studies attempting to develop biological or cognitive markers for detecting the presence or absence of a certain disease; and (iii) prediction of outcomes—studies attempting to predict clinical outcomes in patients with a certain illness, for example, in terms of remission versus relapse. It should also be acknowledged that the majority of the exemplar studies reviewed in the book are taken from the neuroimaging literature. This reflects the fact that, despite notable applications of machine learning methods to other types of data (e.g., genetic risk scores), the overwhelming majority of the studies published so far have used structural or functional neuroimaging data. All of the machine learning methods presented in this book, therefore, are not specific to neuroimaging but can be applied to other types of data.

The book is divided into four parts. In Part I, we introduce the topic of machine learning, review its fundamental concepts, and illustrate its potential applications to brain disorders. In Part II, we provide a more detailed description of some of the most commonly used machine learning algorithms in the literature, as well as some novel and promising approaches that are likely to become popular in the near future. Each chapter includes a discussion of the method and a brief review of relevant examples form the literature. In Part III, we discuss some of the main methodological challenges associated with the application of machine learning to brain disorders. These include, for example, dealing with missing data, high dimensionality, integrating different modalities, heterogeneity, and ethical issues. Finally, in Part IV, we provide a sample code in Python and a step-by-step guide to its implementation; this will enable the reader to practice the application of machine learning to an existing, freely available data set.

Technology companies such as Google, Facebook, and Apple have long been using machine learning methods for big data analysis to predict how people search the Internet, what news they are interested in, and who

their potential friends are. Although the use of machine learning in real-world health care is still in its infancy, its translational application in several areas of medicine including psychiatry and neurology is a matter of when rather than if. We hope that this book will help increase machine learning literacy amongst the next generation of researchers and clinicians, who will be able to use these methods to overcome current research and clinical challenges and, ultimately, generate insights that will improve the life of patients with brain disorders.