Neuroscience and Biobehavioral Reviews 74 (2017) 58–75

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review article

Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications

Sandra Vieira[a,*], Walter H.L. Pinaya[b], Andrea Mechelli[a]

[a] Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, 16 De Crespigny Park, SE5 8AF, United Kingdom
[b] Centre of Mathematics, Computation, and Cognition, Universidade Federal do ABC, Rua Arcturus, Jardim Antares, São Bernardo do Campo, SP CEP 09.606-070, Brazil

ARTICLE INFO

Article history:
Received 2 October 2016
Received in revised form
22 December 2016
Accepted 4 January 2017
Available online 10 January 2017

Keywords:
Deep learning
Machine learning
Neuroimaging
Pattern recognition
Multilayer perceptron
Autoencoders
Convolutional neural networks
Deep belief networks
Psychiatric disorders
Neurologic disorders

ABSTRACT

Deep learning (DL) is a family of machine learning methods that has gained considerable attention in the scientific community, breaking benchmark records in areas such as speech and visual recognition. DL differs from conventional machine learning methods by virtue of its ability to learn the optimal representation from the raw data through consecutive nonlinear transformations, achieving increasingly higher levels of abstraction and complexity. Given its ability to detect abstract and complex patterns, DL has been applied in neuroimaging studies of psychiatric and neurological disorders, which are characterised by subtle and diffuse alterations. Here we introduce the underlying concepts of DL and review studies that have used this approach to classify brain-based disorders. The results of these studies indicate that DL could be a powerful tool in the current search for biomarkers of psychiatric and neurologic disease. We conclude our review by discussing the main promises and challenges of using DL to elucidate brain-based disorders, as well as possible directions for future research.

Contents

* Corresponding author.
E-mail address: sandra.vieira@kcl.ac.uk (S. Vieira).

http://dx.doi.org/10.1016/j.neubiorev.2017.01.002

1. Introduction

In the last two decades, neuroimaging studies of psychiatric and neurological patients have relied on mass-univariate analytical techniques (e.g. statistical parametric mapping). These studies typically compared patients with a diagnosis of interest against disease-free individuals and reported neuroanatomical or neurofunctional differences at group level. The simplicity and interpretability of this approach have led to significant advances in our understanding of the neurobiology of psychiatric and neurological disorders. Mass-univariate analytical techniques, however, suffer from at least two significant limitations. First, statistical inferences are drawn from multiple independent comparisons (i.e. one for each voxel) based on the assumption that different brain regions act independently. This assumption, however, is not in line with our current understanding of brain function in health and disease (Fox et al., 2005; Biswal et al., 2010); for example, several psychiatric and neurological symptoms are best explained by network-level changes in structure and function rather than focal alternations (Mulders et al., 2015; Kennedy and Courchesne, 2008; Sheffield and Barch, 2016). Second, mass-univariate techniques can be used to detect differences between groups but do not allow statistical inferences at the level of the individual. In contrast, a clinician has to make diagnostic and treatment decisions about the person in front of them. These two limitations may have contributed to the limited translational impact of neuroimaging findings in everyday clinical practice so far.

In an attempt to overcome these limitations, the neuroimaging community has developed a growing interest in machine learning (ML), an area of artificial intelligence that aims to develop algorithms that discover trends and patterns in existing data and use this information to make predictions on new data. This is achieved through the use of computational statistics and mathematical optimization (Hastie et al., 2001). ML methods are multivariate and therefore take the inter-correlation between voxels into account, thereby overcoming the first limitation of mass-univariate analytical techniques. In addition, ML methods allow statistical inferences at single subject level and therefore could be used to inform diagnostic and prognostic decisions of individual patients, thereby overcoming the second limitation of mass-univariate analytical techniques (Arbabshirani et al., 2016). ML methods can be divided into two broad categories: supervised and unsupervised learning. In supervised ML, one seeks to develop a function which maps two or more sets of observations to predefined categories or values. In contrast, unsupervised methods seek to determine how the data are organized without using any a priori information supplied by the operator; here the main objective is to discover unknown structure in the data (Hastie et al., 2001).

Over the past decade, several ML methods have been applied to neuroimaging data from psychiatric and neurological patients with varying degrees of success (Arbabshirani et al., 2016; Wolfers et al., 2015). The most popular amongst these methods is Support Vector Machine (SVM), a supervised technique that works by estimating an optimal hyperplane that best separates two classes. When these classes are not linearly separable, SVM uses external functions (kernels) that map the original data into a new feature space where the data become linearly separable (Pereira et al., 2009; Vapnik, 1995). Despite its popularity, SVM has been criticised for not performing well on raw data and requiring the expert use of design techniques to extract the less redundant and more informative features (a step known as "feature selection") (LeCun et al., 2015; Plis et al., 2014). These features, rather than the original data, are then used for classification. While SVM remains a very popular technique within the neuroimaging community, an alternative family of ML methods known as deep learning (DL) (Bengio, 2009) is gaining considerable attention in the wider scientific community (Arbabshirani et al., 2016; Calhoun and Sui, 2016; LeCun et al., 2015). Deep learning methods are a type of representation-learning methods, which means that they can automatically identify the optimal representation from the raw data without requiring prior feature selection. This is achieved through the use of a hierarchical structure with different levels of complexity, which involves the application of consecutive nonlinear transformations to the raw data. These transformations result in increasingly higher levels of abstraction, where higher-level features are more invariant to the noise present in the input data than lower level ones (LeCun et al., 2015). Inspired by how the human brain processes information, the building blocks of DL neural networks – known as "artificial neurons" – are loosely modelled after biological neurons. Artificial neurons are organized in layers. A deep neural network consists of an input layer, two or more hidden layers and an output layer. The input layer comprises the data inputted into the model (e.g. voxel intensity); the hidden layers learn and store increasingly more abstract features of the data; these features are then fed to the output layer that assigns the observations to classes (e.g. controls vs. patients). Learning is achieved through an iterative process of adjustment of the interconnections between the artificial neurons within the network, much like in the human brain (Bengio, 2009). An essential aspect of DL that differentiates it from other ML methods is that the features are not manually engineered; instead, they are learned from the data, resulting in a more objective and less bias-prone process. Besides, the ability to achieve higher orders of abstraction and complexity relative to other ML methods such as SVM makes DL better suited for detecting complex, scattered and subtle patterns in the data (Plis et al., 2014).