NeuroImage 145 (2017) 288–303

Contents lists available at ScienceDirect

NeuroImage

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

“Look at my classifier's result”: Disentangling unresponsive from (minimally) conscious patients

Quentin Noirhomme [a,b,c,*], Ralph Brecheisen [a,b], Damien Lesenfants [d], Georgios Antonopoulos [e], Steven Laureys [e]

[a] Brain Innovation BV, Maastricht, Netherlands
[b] Department of Cognitive Neuroscience, Faculty Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands
[c] Cyclotron Research Centre, University of Liege, Liege, Belgium
[d] School of Engineering and Institute for Brain Science, Brown University, Providence, Rhode Island, USA
[e] Coma Science Group, University Hospital of Liege, Liege, Belgium

A R T I C L E I N F O

Article history:
Accepted 4 December 2015
Available online 12 December 2015

Keywords:
Machine learning
Classifier
Disorders of consciousness
Coma
Vegetative state
Unresponsive wakefulness syndrome
Minimally conscious state
Locked-in syndrome
Diagnosis
Prognosis

A B S T R A C T

Given the fact that clinical bedside examinations can have a high rate of misdiagnosis, machine learning techniques based on neuroimaging and electrophysiological measurements are increasingly being considered for comatose patients and patients with unresponsive wakefulness syndrome, a minimally conscious state or locked-in syndrome. Machine learning techniques have the potential to move from group-level statistical results to personalized predictions in a clinical setting. They have been applied for the purpose of (1) detecting changes in brain activation during functional tasks, equivalent to a behavioral command-following test and (2) estimating signs of consciousness by analyzing measurement data obtained from multiple subjects in resting state. In this review, we provide a comprehensive overview of the literature on both approaches and discuss the translation of present findings to clinical practice. We found that most studies struggle with the difficulty of establishing a reliable behavioral assessment and fluctuations in the patient's levels of arousal. Both these factors affect the training and validation of machine learning methods to a considerable degree. In studies involving more than 50 patients, small to moderate evidence was found for the presence of signs of consciousness or good outcome, where one study even showed strong evidence for good outcome.

Introduction

“Look up. Look down. Squeeze my hand”. These simple commands behaviorally assess the state of consciousness of a patient following a coma. To date, the diagnostic assessment of patients with disorders of consciousness (DOC) is mainly based on the observation of motor and oromotor behavior at the bedside (Giacino et al., 2014). The evaluation of non-reflex behavior, however, is not straightforward because the patient's level of vigilance may fluctuate over time. Also, he or she may suffer from cognitive deficits (e.g., aphasia or apraxia) and/or sensory impairments (e.g., blindness, deafness, paralysis). Reduced, or easily exhausted, motor activity and pain are other factors that may complicate the evaluation. In all these cases, a lack of responsiveness does not necessarily correspond to absence of awareness (Sanders et al., 2012). The identification of unambiguous signs of consciousness in patients with DOC is clinically challenging and of critical importance for establishing a diagnosis, guiding therapeutic decisions and predicting outcome. Therefore, recognizing the subtle difference between unresponsive wakefulness syndrome (UWS) patients (where patients “awaken” from a coma, meaning they open their eyes, but only show reflex behavior, formerly known as vegetative state or apallic syndrome; Laureys et al., 2010) and minimally conscious state (MCS) patients (who show non-reflex movement, e.g., visual fixation or pursuit, localization to pain or following simple commands like “look up” and “squeeze my hand”; Bruno et al., 2011; Giacino et al., 2002) requires repeated evaluations by skilled examiners. Furthermore, it is relatively easy to confuse UWS and locked-in syndrome patients (LIS; Plum and Posner, 1971) who are fully conscious but completely paralyzed except for small movements of the eyes or eyelids. Not surprisingly, up to 40% of patients with UWS are misdiagnosed (Schnakers et al., 2009a). Key elements in the diagnosis are the acquisition of voluntary responses, such as command following, and functional communication which indicates an emergence from UWS (Schnakers et al., 2009a) and MCS, respectively.

Neuroimaging and electrophysiological approaches have been proposed to complement the bedside examination. They offer motor-independent information to improve clinical differentiation and prognostic predictions. Nevertheless, while significant differences have been reported at the group level, most of these results do not allow

* Corresponding author at: Brain Innovation BV Oxfordlaan 55 6229 EV Maastricht The Netherlands.
E-mail address: noirhomme@brainvoyager.com (Q. Noirhomme).

http://dx.doi.org/10.1016/j.neuroimage.2015.12.006
1053-8119/

distinguishing patients at the single-subject level. Some studies, however, extend standard statistical analysis at the single-subject level with expert visual inspection (Stender et al., 2014) or prior hypotheses (for example, Owen et al., 2006; Schnakers et al., 2008; Monti et al., 2010). For example, Owen and colleagues (Owen et al., 2006) instructed a patient to alternate 30-second periods of mental imagery of playing tennis with 30-second periods of rest following a block-design protocol. A single trial consisted of 5 rest vs. imagery cycles. Then, a general linear model contrasting periods of active imagery with periods of rest was computed. Contrasts were constrained by prior hypotheses on activated brain locations; in this case, the supplementary motor area**.** Significant activation in the predefined brain locations is then used as indication that the patient is correctly performing the task. Calculating and thresholding a single variable or group of variables has also been proposed. For example, spectral entropy summarizes EEG signals as a single value which can distinguish patients in acute state with good accuracy (Gosseries et al., 2011).

Machine learning techniques have the potential to make more effective use of neuroimaging and electrophysiological data and allow diagnosis and prognosis at the single-subject level. Instead of considering features/activations univariately, they combine information in a multivariate way which allows them to highlight differences that might otherwise remain undetected. They are also not biased by prior hypotheses on location or time because they do not focus on the detection of a specific activation pattern but rather on a data-driven estimation of the most discriminative pattern within a trial or class. This can be an advantage given that prior hypotheses may no longer hold in pathological situations. It has been shown that data obtained from patients often exhibits higher inter-trial as well as inter-individual variability than data obtained from controls (Goldfine et al., 2011; King et al., 2013a; Lulé et al., 2013). Machine learning techniques provide a way to quantify differences in neural responses at the level of the single patient. Also, their statistical validation is limited to a single test which is independent of the number of features. This has the added advantage of also limiting the multiple comparison problem.

Until now, machine learning techniques have been applied to individual diagnosis using two main approaches: (1) detection of command-following and (2) prediction of diagnosis and outcome using structural or functional data. The first approach uses data from only a single subject measured over time and has its origin in brain-computer interface (BCI) research. The goal is to assess whether a subject is capable of following commands by measuring his/her brain activity during a functional task. For example, in the tennis paradigm mentioned above, the level of activation in each gray matter voxel could be averaged over a short time period. The average activations of all gray matter voxels could then be used as input features to train a classifier that detects the transition between rest and active imagery states. If the classification accuracy exceeds a given threshold, the subject can be considered to be able to correctly modulate his/her brain activity according to the given commands. This would be equivalent to behavioral command-following, which is a sign of MCS.

The second approach uses data obtained from multiple subjects and tries to derive a prediction model that can be used on individual subjects. For example, resting-state fMRI data might be acquired from a group of patients and healthy controls after which connectivity matrices of certain resting-state networks are calculated for all subjects. Since each group of subjects is likely to have a specific pattern of fMRI connectivity a classifier can be trained which uses the connectivity features to distinguish between the groups, for example, controls versus unresponsive patients. If classification accuracy is high enough the resulting model can then be used to classify (diagnose) new patients. Instead of resting-state networks, features can also be derived from EEG, structural MRI, diffusion tensor imaging (DTI) or positron emission tomography (PET).

In this paper, we present a survey of the literature on the use of electrophysiology and neuroimaging for diagnosing patients with disorders of consciousness (DOC). We compare machine learning techniques with studies based on univariate analysis and simple thresholding. We first provide a brief introduction of the diagnosis of DOC and list key points to take into account when machine learning techniques are used to improve the diagnosis in a clinically useful way. We will then give an overview of previous work done in the two areas mentioned earlier: (1) detection of command-following and (2) prediction of diagnosis and outcome based on multi-subject data. We will highlight the main limitations common to many studies and offer a number of suggestions for further investigation. Finally, we discuss several challenges which the field needs to overcome in order to translate machine learning techniques into clinical practice.

Machine learning for diagnosis of disorders of consciousness

Current practice in diagnosing DOC

Disorders of consciousness are currently mostly based on consensus diagnosis or using the Coma Recovery Scale-Revised (CRS-R; Giacino et al., 2004). Consensus diagnosis is based on behavioral observations of caregivers and is the most common type of diagnostic procedure in non-specialized centers. These centers would likely benefit the most from an automated diagnostic procedure given the fact that they do not usually employ DOC specialists. The rate of misdiagnosis of UWS patients by clinical consensus methods is up to 40% (Andrews et al., 1996; Childs and Mercer, 1996; Schnakers et al., 2009a). However, this error rate can be reduced by using standardized scoring systems such as the CRS-R, which is currently the most validated and sensitive method for behavioral discrimination of patients with DOC. Diagnosis remains challenging, however, because patients typically show considerable fluctuations in the level of consciousness or arousal over time. The examiner may obtain clear evidence of volitional behavior during one examination but fail to do so in another examination conducted hours or even minutes later (Giacino et al., 2014). For this reason, repeated CRS-R assessments performed by trained and experienced caregivers are essential to establish a reliable final diagnosis (Giacino et al., 2004). Repeating the assessment at least 5 times within a short period (e.g., 2 weeks) has been shown to be most accurate for establishing a diagnosis in chronic DOC patients (Wannez et al., 2016).

Despite the fact that the repeated CRS-R assessment is becoming a standard for diagnosing DOC there is still a chance that patients are incorrectly diagnosed as behaviorally unresponsive. Neuroimaging studies have shown that up to 20% of behaviorally unresponsive patients still show signs of awareness based on their brain activations (see Table 1, specificity). This is one reason why some diagnoses depend on the outcome of imaging or electrophysiological experiments. For example, functional locked-in syndrome patients show extreme behavioral motor dysfunction but still have preserved higher cognitive functions as measured by functional imaging (Bruno et al., 2011). Also, the results of repeated CRS-R assessments are reported in varying ways. Some report the patient's consciousness state only on the day of data recording while others only report the highest consciousness state measured across the multiple assessments. Ideally, results from the CRS-R assessment at the time of data recording and results of any repeated assessments should be reported together to give the clearest picture of a patient's consciousness state over time.

Challenges and limitations of current practice

Diagnosing disorders of consciousness is a challenging problem for a number of reasons which we will discuss shortly. These challenges will also affect any machine learning methods applied to the data.

Lack of gold standard

Difficulty in establishing a reliable behavioral diagnosis of DOC, as mentioned briefly before, is one of the main reasons for a lack of