Neuroscience and Biobehavioral Reviews 80 (2017) 538–554

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

The impact of machine learning techniques in the study of bipolar disorder: A systematic review

Diego Librenza-Garcia[a,b], Bruno Jaskulski Kotzian[b], Jessica Yang[c], Benson Mwangi[d], Bo Cao[d], Luiza Nunes Pereira Lima[b], Mariane Bagatin Bermudez[a,b], Manuela Vianna Boeira[b], Flávio Kapczinski[e], Ives Cavalcante Passos[b,f,*]

[a] Graduation Program in Psychiatry, Universidade Federal das Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brazil
[b] Bipolar Disorder Program, Laboratory of Molecular Psychiatry, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, RS, 90035-903, Brazil
[c] University of Texas at Austin College of Pharmacy, United States
[d] Department of Psychiatry and Behavioral Sciences, The University of Texas Science Center at Houston, Houston, TX, USA
[e] McMaster's Department of Psychiatry and Behavioral Neurosciences, Hamilton, Canada
[f] Graduation Program in Psychiatry and Department of Psychiatry, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, 90035-903, Brazil

ARTICLE INFO

Keywords:
Machine learning
Big data
Bipolar disorder
Suicide
Diagnosis
Support vector machine
Prediction
Predictive analysis
Pattern recognition
Neuroimaging

ABSTRACT

Machine learning techniques provide new methods to predict diagnosis and clinical outcomes at an individual level. We aim to review the existing literature on the use of machine learning techniques in the assessment of subjects with bipolar disorder. We systematically searched PubMed, Embase and Web of Science for articles published in any language up to January 2017. We found 757 abstracts and included 51 studies in our review. Most of the included studies used multiple levels of biological data to distinguish the diagnosis of bipolar disorder from other psychiatric disorders or healthy controls. We also found studies that assessed the prediction of clinical outcomes and studies using unsupervised machine learning to build more consistent clinical phenotypes of bipolar disorder. We concluded that given the clinical heterogeneity of samples of patients with BD, machine learning techniques may provide clinicians and researchers with important insights in fields such as diagnosis, personalized treatment and prognosis orientation.

1. Introduction

Bipolar disorder affects about 2% of the world's population with sub-threshold forms affecting an additional 2% of the population (Merikangas et al., 2007). According to the World Health Organization, bipolar disorder is among the 10 leading causes of disability-adjusted life years in young adults (Mathers et al., 2006). Rates of completed suicide in patients with bipolar disorder are 7.8% in men and 4.9% in women (Nordentoft, 2011), and life expectancy has been reported to decrease by 9 years in patients with bipolar disorder (Crump et al., 2013). Although several types of interventions may be used in order to prevent and treat mood episodes, they are frequently suboptimal, and about 60% of the patients relapse into depression or mania within two years of treatment initiation (Gitlin et al., 1995). In addition, current approaches to diagnosing bipolar disorder may not be completely effective, having an average delay of ten years between the first symptoms and the formal diagnosis (Lish et al., 1994). This framework illustrates important challenges in the current treatment approaches, diagnosis, and prevention in bipolar disorder.

Evidence-based medicine has helped us understand risk factors, optimal treatments and prognosis of bipolar disorder by using traditional statistical methods which primarily provide average group-level results (Sackett et al., 1996). However, in a recent article, Greenhalg and colleagues have called our attention to the fact that some statistically significant results may not represent a real benefit for an individual patient and that subjects in clinical trials may not always reflect the multimorbidity profile of real life patients (Greenhalgh et al., 2014). This may be particularly true in the field of bipolar disorder, where clinical heterogeneity is a very important factor. In light of these findings, techniques that aim at developing tailor-made psychiatric care to the individual, such as machine learning, have been gaining ground in psychiatric research (Huys et al., 2016).

Big data is a broad term used to denote volumes of large and complex measurements, as well as the velocity that data is created.

* Corresponding author at: Federal University of Rio Grande do Sul, Avenida Ramiro Barcelos, 2350, Zip Code: 90035-903, Porto Alegre, RS, Brazil.
E-mail addresses: diegolibrenzagarcia@gmail.com (D. Librenza-Garcia), brunokotzian@hotmail.com (B.J. Kotzian), jessica.yang10@gmail.com (J. Yang), benson.mwangi@gmail.com (B. Mwangi), cloudbocao@gmail.com (B. Cao), luiza.npl@gmail.com (L.N. Pereira Lima), mari.bermudez@yahoo.com.br (M.B. Bermudez), manuelaboeira@yahoo.co.uk (M.V. Boeira), flavio.kapczinski@gmail.com (F. Kapczinski), ivescp@yahoo.com.br (I.C. Passos).

http://dx.doi.org/10.1016/j.neubiorev.2017.07.004
Received 10 February 2017; Received in revised form 15 June 2017; Accepted 8 July 2017
Available online 18 July 2017
0149-7634/

Another crucial characteristic of big data is the variety of levels in which data is created, from the molecular level, including genomics, proteomics and metabolomics, to clinical, sociodemographic, administrative, environmental, and even social media information (Passos et al., 2016b). In order to analyze big data, several machine learning methods (also known as pattern recognition techniques) have been developed during the last years. In short, one may say that first the algorithm analyses a 'training' dataset to establish a function able to distinguish individual subjects across groups. Once that has been done, the model can be applied to a new dataset, and the accuracy of the method can be measured in this new scenario. Later improvements of the model can be acquired, either by changing the algorithm or by performing additional feature reduction in the dataset (Lantz, 2015). As a result, these algorithms are ideal for assessing multifactorial disorders, and to estimate the probability of a particular outcome at an individual level (Mwangi et al., 2012).

The present study aims to review data in which bipolar disorder patients were assessed by using machine learning techniques regarding different outcomes. Our focus was mainly on studies that assessed diagnosis. However, we also included studies that assessed treatment, prognosis and development of data-driven phenotypes. We also provided a brief explanation about the most relevant principles of machine learning and its limitations in the supplementary material, since these techniques are relatively new in the field of psychiatry. Our overarching goal was to show how these new techniques are likely to support important clinical decisions in the forthcoming years.

2. Methods

2.1. Search strategy

We searched PubMed, Embase and Web of Science for articles published between January 1960, and January 2017 by using the following keywords: ("Big data" OR "Artificial Intelligence" OR "Machine Learning" OR "Gaussian process" OR "Cross-validation" OR "Cross validation" OR "Crossvalidation" OR "Regularized logistic" OR "Linear discriminant analysis" OR "LDA" OR "Random forest" OR "Naïve Bayes" OR "Least Absolute selection shrinkage operator" OR "elastic net" OR "LASSO" OR "RVM" OR "relevance vector machine" OR "pattern recognition" OR "Computational Intelligence" OR "Computational Intelligences" OR "Machine Intelligence" OR "Knowledge Representation" OR "Knowledge Representations" OR "support vector" OR "SVM" OR "Pattern classification") AND ("Bipolar Disorder" OR "Bipolar Disorders" OR "Manic-Depressive Psychosis" OR "Manic Depressive Psychosis" OR "Bipolar Affective Psychosis" OR "Manic-Depressive Psychoses" OR "Mania" OR " OR "Manic State" OR "Manic States" OR "Bipolar Depression" OR "Manic Disorder" OR "Manic Disorders" OR "Bipolar euthymic"). We also searched the reference lists to find potential articles to include. There were no language restrictions.

2.2. Eligibility criteria

This systematic review was performed according to the PRISMA statement (Liberati et al., 2009). Articles met the inclusion criteria if they assessed bipolar disorder patients using machine learning techniques. Technical and theoretical studies that used machine learning techniques but did not assess bipolar disorder patients were excluded. We also excluded studies that included only individuals below 18 years of age.

2.3. Data collection, extraction and statistical analysis

Two researchers (DLG and LNPL) independently screened titles and abstracts of the identified articles. They also obtained and read the full texts of potential articles, supervised by ICP who made the final decision in cases of disagreement. Data extracted from the articles included year of study publication, data used in the machine learning model (i.e., neuroimaging, blood biomarkers, clinical and demographical characteristics, among others), sample size, diagnoses assessed in the study, machine learning algorithm, and statistical measure of performance (i.e., accuracy, sensitivity, specificity, area under the curve, true positive, false positive, true negative and false negative). When this data was not available, we requested it from the authors.

We performed a meta-analysis of diagnostic accuracy with the classification studies. For this analysis, we included studies that used neuroimaging data (either structural or functional) to assess patients with bipolar disorder compared with healthy controls. Articles that assessed patients with bipolar disorder compared to patients with other psychiatric diagnosis were excluded. We also excluded studies performed in subjects with less than 15-year-old. We used the package *mada* from R (version number 3.3.1) to perform the analysis and to build the Summary Receiver Operating Characteristic (SROC) curve of sensitivity and specificity, as previously described (Reitsma et al., 2005).

3. Results and discussion

We found 757 potential abstracts and included 51 articles in the present review, being one of them added after reference screening (Fig. 1). A list of the included articles as well as its most relevant characteristics and findings is presented in Table 1 (classification studies) and tables S1 and S2 (clinical outcomes prediction and unsupervised learning studies). Most of the studies focused on diagnostic classification (38 studies) in order to distinguish bipolar disorder from schizophrenia, unipolar depression, healthy controls and other conditions. Of these, 11 used structural neuroimaging (Besga et al., 2012; Chen et al., 2014; Fung et al., 2015; Hajek et al., 2015; Koutsouleris et al., 2015; Mwangi et al., 2016; Redlich et al., 2014; Rocha-Rego et al., 2014; Sacchet et al., 2015; Schnack et al., 2014; Serpa et al., 2014), 13 used functional neuroimaging (Almeida et al., 2013; Anticevic et al., 2014; Arribas et al., 2010; Costafreda et al., 2011; Du et al., 2015; Frangou et al., 2017; Grotegerd et al., 2014, 2013; Jie et al., 2015; Kaufmann et al., 2017; Mourão-Miranda et al., 2012; Rive et al., 2016; Roberts et al., 2016), 5 used genetic analysis (Acikel et al., 2016; Chuang and Kuo, 2017; Dmitrzak-Weglarz et al., 2015; Pirooznia et al., 2012; Struyf et al., 2008), 4 used electroencephalographic measures (Erguzel et al., 2016, 2015; Johannesen et al., 2013; Khodayari-Rostamabad et al., 2010), 3 used neuropsychological tests, either alone or coupled with clinical observations and serum biomarkers (Akinci et al., 2013; Besga et al., 2015; Wu et al., 2016b), and 2 used a panel of serum biomarkers (Haenisch et al., 2016; Pinto et al., 2017). A total of 7 studies focused on predicting clinical outcomes, such as depression relapse and suicide (Salvini et al., 2015), mood changes (Faurholt-Jepsen et al., 2016; Gentili et al., 2017; Valenza et al., 2014, 2013) and suicide (Levey et al., 2016; Niculescu et al., 2015; Passos et al., 2016a). We found only 2 articles predicting treatment response or adverse effects (Castro et al., 2016; Wade et al., 2016), and 4 studies that used unsupervised or semi-supervised machine learning to identify homogeneous groups of patients (Bansal et al., 2012; Hall et al., 2012; Wahlund et al., 1998; Wu et al., 2016b).

3.1. Classification studies

Bipolar disorder particularly illustrates the dilemma of diagnostic systems solely based on clinical judgment, which may lead to misdiagnosis or treatment delay. It is known that bipolar disorder has an average delay of ten years between the first symptoms and formal diagnosis (Lish et al., 1994). It is also known that only 20% of patients with bipolar disorder who are experiencing a depressive episode are diagnosed with bipolar disorder within the first year of seeking treatment (Goldberg et al., 2001).