Epilepsy & Behavior xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Epilepsy & Behavior

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

Machine learning as a new paradigm for characterizing localization and lateralization of neuropsychological test data in temporal lobe epilepsy

Brandon Frank [a], Landon Hurley [a], Travis M. Scott [a], Pat Olsen [a], Patricia Dugan [b], William B. Barr [b,*]

[a] Department of Psychology, Fordham University, 441 East Fordham Road, Bronx, NY 10458, United States of America
[b] Department of Neurology, NYU School of Medicine, New York, NY 10016, United States of America

ARTICLE INFO

Article history:
Received 25 May 2018
Revised 5 July 2018
Accepted 5 July 2018
Available online xxxx

Keywords:
Temporal lobe epilepsy
Machine learning
Neuropsychology
Localization
Lateralization

ABSTRACT

In this study, we employed a kernel support vector machine to predict epilepsy localization and lateralization for patients with a diagnosis of epilepsy ($n = 228$). We assessed the accuracy to which indices of verbal memory, visual memory, verbal fluency, and naming would localize and lateralize seizure focus in comparison to standard electroencephalogram (EEG). Classification accuracy was defined as models that produced the least cross-validated error (CVϵ). In addition, we assessed whether the inclusion of norm-based standard scores, demographics, and emotional functioning data would reduce CVϵ. Finally, we obtained class probabilities (i.e., the probability of a particular classification for each case) and produced receiver operating characteristic (ROC) curves for the primary analyses. We obtained the least error assessing localization data with the Gaussian radial basis kernel function (RBF; support vectors = 157, CVϵ = 0.22). There was no overlap between the localization and lateralization models, such that the poorest localization model (the hyperbolic tangent kernel function; support vectors = 91, CVϵ = 0.36) outperformed the strongest lateralization model (RBF; support vectors = 201, CVϵ = 0.39). Contrary to our hypothesis, the addition of norm, demographics, and emotional functioning data did not improve the accuracy of the models. Receiver operating characteristic curves suggested clinical utility in classifying epilepsy lateralization and localization using neuropsychological indicators, albeit with better discrimination for localizing determinations. This study adds to the existing literature by employing an analytic technique with inherent advantages in generalizability when compared to traditional single-sample, not cross-validated models. In the future, class probabilities extracted from these and similar analyses could supplement neuropsychological practice by offering a quantitative guide to clinical judgements.

1. Introduction

Neuropsychological batteries serve multiple important functions in epilepsy treatment with one of the major functions being localizing/lateralizing functional impairment associated with the seizure focus [1]. The conclusions made by neuropsychologists in the analysis of neuropsychological data often demand complex rather than simplistic categorization. Novel analytic techniques such as machine learning, which employ complex categorization functions, may contribute to neuropsychological research and practice in a manner superior to traditional analyses. In particular, these analyses may provide an optimal quantitative guide to assist clinical judgements.

The vast majority of neuropsychological investigations regarding localization and lateralization in epilepsy employ linear models. Millis [2] and Schatz et al. [3] detailed statistical violations that have become common practice in these studies. Failure to properly account for statistical assumptions may lead to substantial inaccuracy [4]. In addition, statistical adjustments made for multiple comparisons (e.g., the Bonferroni adjustment and false discovery rate) demand careful consideration because of the strengths and limitations of each technique [2].

Alternately, little has been published using algorithmic approaches with neuropsychological data (see [5–7] for exceptions). However, advanced methodologies have shown promise in assisting clinical judgements. Bowden and Loring [8] proposed multiple-level likelihood ratios as a potential alternative to dichotomous sensitivity and specificity analyses. We suggest that machine learning provides a useful analytic paradigm. Most importantly, these algorithms prioritize accuracy, which remains of upmost clinical concern, over the interpretation of individual predictors.

1.1. Machine learning

In general, classification involves the development of numerical models that sort data into two or more groups. Classification algorithms use predictor variables to sort data points according to a preset rule. For

* Corresponding author at: NYU Langone Comprehensive Epilepsy Center, 223 East 34th Street, New York, NY 10016, United States of America.
E-mail address: william.barr@nyumc.org (W.B. Barr).

https://doi.org/10.1016/j.yebeh.2018.07.006
1525-5050/

example, in support vector machines (SVMs), error is permitted up to a threshold of magnitude [9]. However, classification fails to be useful when models fit a particular dataset but do not generalize to other relevant data. This occurs in regression models [2] and algorithms [10] that are attuned to random noise in the data rather than parsimonious and meaningful depictions of variation (e.g., overfitting). Machine learning contends with this problem by dividing data into training sets and testing sets. Algorithms produce a model using observed variables to optimally predict class labels and then cross-validate that model with testing sets. This method possesses inherent advantages in generalizability over single-sample models [9,11].

Kernel SVMs classify using a fixed nonlinear mapping of the data into high-dimensional space, such that a linear classification becomes tenable [11,12]. In other words, kernel SVMs iteratively test transformations, interactions, and other functions on the data to optimize classification [13,14]. As mentioned, these multifaceted procedures may be well-suited to the complex categorization tasks necessary in neuropsychology. Machine learning algorithms possess the capacity to minimize cross-validated error ($CV\epsilon$), even in the presence of complex and noisy data [9]. Kernel SVMs provide clear results from a complex and unspecified process. Machine learning algorithms will not statistically assess the relative contribution of each predictor variable, but they will comprehensively use the data to optimize classification.

In this retrospective study, our aim was to determine whether machine learning algorithms would distinguish localized temporal lobe epilepsy (TLE) from other forms of focal epilepsy using neuropsychological data. Given previous research [15–21], we hypothesized that SVM algorithms would classify seizure localization with greater accuracy than seizure lateralization. Furthermore, we anticipated that lateral classifications would suggest little clinical utility. This was defined as predictive accuracy slightly greater than chance. Second, we hypothesized that both localization and lateralization would be more accurately classified with the inclusion of norm-based standard scores, demographics, and emotional functioning data. This study has the potential to inform researchers regarding new methodologies in the characterization of epilepsy. In addition, future practitioners could employ case probabilities extracted from these algorithms as a guide for clinical judgements.

2. Methods

2.1. Participants

This was a retrospective analysis of 228 patients who completed neuropsychological testing as part of a comprehensive neurological evaluation (Table 1). All participants were assessed by the NYU-CEC from 2003 to 2008. In addition to neuropsychological data, inclusion criteria included a confirmed epilepsy diagnosis through either routine electroencephalogram (EEG) or video-electroencephalogram (VEEG) as indicated in electronic or hard copy medical chart (see Fig. 1). All eligible batteries were given in English. Batteries given in Spanish ($n = 30$) or another language ($n = 4$) had either missing data or an unconfirmed diagnosis. Participants were coded for epilepsy lateralization with confirmed right versus left seizure focus. In addition, we coded participants as having TLE only, versus all other specified epilepsy subtypes (i.e., frontal, frontotemporal, and other focal). Nonlocalized cases of epilepsy were not included for this analysis.

We recognize that epilepsy subtypes are nonequivalent. However, we created the 'extratemporal' group for several purposes. First, we wanted to determine the accuracy in which we could categorize TLE against all other groups. We feel that this prediction serves an important clinical purpose (e.g., prior to hippocampectomy or temporal lobectomy). Second, heterogeneity in the extratemporal group should, if anything, limit predictive accuracy (against our hypothesis). Third, we wanted to compare the accuracy of lateralization and localization categorizations. In order for this comparison to be meaningful in a machine learning context, both analyses would need to be dichotomous. It should be noted that machine learning analyses could be run to distinguish localized versus generalized epilepsy, or to predict classes across epilepsy subtypes. However, this was not done in the current study (see Discussion).

For patients with multiple neuropsychological test sessions, only data from the earliest testing session were included. Exclusion criteria included: 1) a diagnosis of nonepileptic seizures, 2) prior resective brain surgery, and 3) additional neurological disorders that would complicate the etiology of the patient's cognitive difficulties (e.g., history of moderate to severe traumatic brain injury (TBI), substantial stroke outside of the surgical area, and similar neurological disorders). All study procedures were approved from NYU School of Medicine and Fordham University human subjects committee review boards.

Table 1
Patient demographic information.

	Patients ($n = 228$)
Age in years (M, SD)	36.9 (12.8)
Sex (F/M)	125/103
Years of education (M, SD)	14.7 (2.6)
Handedness	Right-handed: 185 (81%)
	Left-handed: 28 (12%)
	Ambidextrous: 4 (2%)
	Unknown: 11 (5%)
Seizure localization	Right temporal: 63 (28%)
	Left temporal: 93 (41%)
	Bilateral temporal: 12 (5%)
	Frontal: 20 (9%)
	Frontotemporal: 20 (9%)
	Other focal: 10 (4%)
	Unknown: 10 (4%)

2.2. Measures

Neuropsychological testing consisted of a 4-hour battery designed to aid in the localization of the seizure focus. This battery included measures of learning and memory, visuospatial functioning, and executive functioning. Additional measures of subjective mood complaints (i.e., depression, anxiety) and quality of life were also administered.

2.2.1. California Verbal Learning Test (CVLT)

For verbal list learning, the present study used both the CVLT and CVLT-II because NYU-CEC transitioned from the original CVLT to the CVLT-II during the current study period. From a total of 223 participants with CVLT data, 128 had CVLT data and 95 had CVLT-II data. The CVLT is a serial word learning task with five learning trials, a single presentation of a second list with recall, followed by free recall of the original list [22]. After a delay of approximately 20 min, free and cued recalls and recognition are tested. The CVLT contains 16 words from four semantic categories (i.e., spices and herbs, fruits, tools, and clothing), and CVLT-II contains 16 words from the following semantic categories: animals, vegetables, modes of transportation, and furniture [23]. California Verbal Learning Test norm-based standardized scores were used as generated by the computer scoring program [24].

2.2.2. Wechsler Memory Scale — III logical memory (WMS-III LM)

WMS-III logical memory I and II is a story memory task where two short stories are presented orally [25]. The examinee is asked to retell each story from memory immediately. After a delay of 20–30 min, long-term narrative memory is assessed with free recall and recognition tasks in which the examinee is asked to retell both stories from the immediate condition, then asked yes/no questions about both stories. We used the age-adjusted scaled scores (M = 10, SD = 3) that are in the published manual.