Journal of Neuroscience Methods 212 (2013) 308–316

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

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

Clinical Neuroscience

Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis

Jonathan J. Halford[a,*], Robert J. Schalkoff[b], Jing Zhou[b], Selim R. Benbadis[c], William O. Tatum[d], Robert P. Turner[a], Saurabh R. Sinha[e], Nathan B. Fountain[f], Amir Arain[g], Paul B. Pritchard[a], Ekrem Kutluay[a], Gabriel Martz[a], Jonathan C. Edwards[a], Chad Waters[h], Brian C. Dean[h]

[a] Department of Neurosciences, Medical University of South Carolina, Charleston, SC, United States
[b] Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, United States
[c] Department of Neurology, University of South Florida & Tampa General Hospital, Tampa, FL, United States
[d] Department of Neurology, Mayo Clinic Jacksonville, Jacksonville, FL, United States
[e] Department of Medicine, Division of Neurology, Duke University Medical Center, Durham, NC, United States
[f] Department of Neurology, University of Virginia, Charlottesville, VA, United States
[g] Department of Neurology, Vanderbilt University Medical Center, Nashville, TN, United States
[h] School of Computing, Clemson University, Clemson, SC, United States

HIGHLIGHTS

- We describe our new web-based software for collecting expert opinion on paroxysmal activity in routine scalp EEG.
- We report that inter-rater correlation among our groups of 11 board-certified EEG scorers was only moderate.
- Our machine learning analysis suggests that our EEG database needs to be larger than its current size to adequately represent the variability of waveform morphologies in EEG.
- Our artificial neural network machine learning classifiers performed better than our Bayesian classifiers and the wavelet features were the most useful.

ARTICLE INFO

Article history:
Received 20 October 2011
Received in revised form 6 November 2012
Accepted 8 November 2012

Keywords:
Electroencephalography
Electroencephalogram (EEG)
Spike detection
Epileptiform transient
Automated interpretation
Computerized interpretation

ABSTRACT

The routine scalp electroencephalogram (rsEEG) is the most common clinical neurophysiology procedure. The most important role of rsEEG is to detect evidence of epilepsy, in the form of epileptiform transients (ETs), also known as spike or sharp wave discharges. Due to the wide variety of morphologies of ETs and their similarity to artifacts and waves that are part of the normal background activity, the task of ET detection is difficult and mistakes are frequently made. The development of reliable computerized detection of ETs in the EEG could assist physicians in interpreting rsEEGs. We report progress in developing a standardized database for testing and training ET detection algorithms. We describe a new version of our EEGnet software system for collecting expert opinion on EEG datasets, a completely web-browser based system. We report results of EEG scoring from a group of 11 board-certified academic clinical neurophysiologists who annotated 30-s excepts from rsEEG recordings from 100 different patients. The scorers had moderate inter-scorer reliability and low to moderate intra-scorer reliability. In order to measure the optimal size of this standardized rsEEG database, we used machine learning models to classify paroxysmal EEG activity in our database into ET and non-ET classes. Based on our results, it appears that our database will need to be larger than its current size. Also, our non-parametric classifier, an artificial neural network, performed better than our parametric Bayesian classifier. Of our feature sets, the wavelet feature set proved most useful for classification.

Published by Elsevier B.V.

1. Introduction

The routine scalp electroencephalogram (rsEEG) is the most common clinical neurophysiology procedure. It consists of a 20–30 min recording from approximately 20 scalp electrodes. Over one million outpatient rsEEGs and over 50,000 inpatient rsEEGs are performed in the United States every year based on Centers for

* Corresponding author at: 96, Jonathan Lucas Street, Suite 307 CSB, Charleston, SC 29425, United States. Tel.: +1 843 792 3223; fax: +1 843 792 8626.
E-mail address: halfordj@musc.edu (J.J. Halford).

0165-0270/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jneumeth.2012.11.005

Medicare and Medicaid Services (CMS) and the National Inpatient Survey in 2007. EEGs are mainly performed to detect evidence of epilepsy.

The EEG of a patient with epilepsy is characterized by occasional epileptiform transients (ETs) consisting of spikes of 20–70 ms and sharp waves of 70–200 ms in duration. Detecting ETs in EEG is important because their presence is predictive of recurrence in patients following a first seizure (Van Donselaar et al., 1992) and is useful in supporting a diagnosis of epilepsy (Fountain and Freeman, 2006). However, due to the wide variety of morphologies of ETs and their similarity to waves that are part of the normal background activity and to artifacts (i.e. extracerebral potentials from muscle, eyes, heart, electrodes, etc.), the detection of ETs is far from straightforward.

Despite the importance of rsEEG, little has changed in the typical way that rsEEG is recorded and interpreted over the last 25 years with the exception of the introduction of digital acquisition, digital display, and digital montage reformatting. Currently, an electroencephalographer (EEGer) detects ETs in rsEEG by visual inspection of 10–20 s of the EEG signals at a time. These EEG signals are generally unprocessed with the exception of rudimentary digital filtering. It is well-known that rsEEGs are frequently misinterpreted by neurologists without neurophysiology fellowship training (who are the majority of neurologists interpreting rsEEG) (Benbadis, 2007). Misinterpretation of the rsEEG can adversely affect patients, leading to the misdiagnosis of epilepsy and the inappropriate use of antiepileptic medications for many years or decades as well as delay in the treatment of the true underlying cause of the seizure-like events, which can be cardiac arrhythmias, psychogenic events, or other types (Benbadis, 2007). Patients with non-epileptic events endure an average of 7 years of antiepileptic medication exposure before they are correctly diagnosed (Reuber et al., 2002).

Despite the problem of rsEEG misinterpretation, there are no commonly accepted ET-detection software programs or devices available. Few studies have measured the performance of automated ET-detection software and almost none have compared the performance between software programs (Halford, 2009). Based on the few studies which have been done and based on the commonly held opinion of academic EEG interpreters, current ET-detection software has an insufficient sensitivity, requiring a neurologist to look over each EEG to make sure that ETs are not missed (Ver Hoef et al., 2010). Current software also has an insufficient specificity, requiring a neurologist to look through a large list of detections, many of which are false positive detections (De Lucia et al., 2008; Halford, 2009; Indiradevi et al., 2008). Accurate automated ET-detection software could improve patient care by improving the accuracy of rsEEG interpretation by neurologists with limited training in EEG interpretation. Additionally, accurate ET-detection software could increase the speed and accuracy of interpreting prolonged outpatient EEG recordings. Prolonged outpatient EEG recordings (lasting at least several hours) will likely become more common in the future since recent studies have demonstrated that recordings longer than the typical 20–25 min rsEEG have an improved yield in detecting ETs (Losey and Uber-Zak, 2008; Modur and Rigdon, 2008). ET detection software can also be applied to prolonged EEG recordings made in inpatient settings such as epilepsy monitoring units or intensive care units (assuming that these ET detection algorithms were trained specifically for inpatient EEG monitoring).

Because of the progressive advances in (1) the processing power of computing systems, (2) the sophistication of machine learning (ML) model creation and (3) the ongoing expansion of proprietary (non-freely available) testing and training datasets for ET detection, EEG automated ET detection software will likely attain the sensitivity and specificity of the average academic clinical neurophysiologist for the detection of ETs in rsEEG within the next decade, if they have not already. But this medical advance will be slowed in its translation into medical practice if clinicians do not trust that commercially available ET detection software performs as well as advertised. Savy physicians, who are trained in evidence-based medicine, will not completely trust their own assessment of these ET detection algorithms. (One of the basic tenets in evidence-based medicine is that trusting one's own clinical experience in the use of a drug, medical tool or device is like trusting in a study of non-randomized, non-controlled data from one observer with a low number of observations.) Therefore, a standardized database (including a freely available training database and a non-freely available testing database) of rsEEG containing ETs and artifacts with sufficient expert opinion and funded by a federal or foundation agency is needed to provide independent confirmation of the performance of commercially available ET detection software and to help set minimum performance standards. We have made it our goal to create this standardized database, and this paper reports our initial efforts.

Since there is considerable variability in the waveform morphologies of both ETs and artifacts within rsEEG, this standardized database will probably need to be quite large. Two central challenges to creating this database are (1) streamlining the process of acquiring expert opinion on rsEEG data to hold down the cost of the project and (2) measuring when the database has become sufficiently large to accurately measure the performance of ET-detection algorithms as applied in clinical neurophysiology practice. In this paper, we describe tools we have created to help us address these challenges and our preliminary data analysis.

To address the first problem (to streamline the process of collecting expert opinion), we created EEGnet, a web-based system for scoring scalp EEG recordings. Past methods for the collection of expert opinion on EEG have been time-consuming and logistically challenging since EEG recordings had to be mailed out on CDs (or other media) and annotations tabulated by hand. More recent methods include installing EEG review software locally and creating annotations within EEG recordings which are stored locally and then tabulated by hand. Collecting expert opinion in the 21st century is now easier due to advances in computer technology, specifically web-based software applications. Our EEGnet system presents EEG data in a familiar visual format within a web browser to clinical EEG interpreters and allows efficient annotation. Our method of displaying EEG within a web browser (coded in JavaScript) with no local install and no significant local EEG data caching is unique, to our knowledge. We have also developed a new two-phase method for collecting expert opinion on EEG events, which we have reported previously (Halford, 2010; Halford et al., 2011). In the first phase of this method, expert interpreters mark *all* paroxysmal EEG events in a recording and in the second phase these events are categorized. This scoring method allows us to quantify true negative ET detections and to calculate not only the sensitivity but also the specificity metric of ET-detection algorithm performance.

To address the second problem (determining the proper size of a standardized database to test/train automated detection algorithms), we have created (ML) models to perform categorization of transient EEG events in order to try to detect when the sufficient size of the database has been reached. The question of how large a standardized database needs to be in order to train and test detection algorithms is important. If a standardized database is too small, detection algorithms trained using a database will not perform well when applied in clinical practice. This is a common outcome in the field of automated ET-detection (Halford, 2009). If a standardized database is created too large, excessive effort and funds will have been spent to create the database. This is a less common, but possible outcome. In order to measure if our standardized EEG dataset is sufficiently large to accurately predict the performance