Knowledge-Based Systems 105 (2016) 48–59

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Automatic diagnosis of alcohol use disorder using EEG features

Wajid Mumtaz[a], Pham Lam Vuong[a], Likun Xia[c], Aamir Saeed Malik[a,*], Rusdi Bin Abd Rashid[b]

[a] Center for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, University Technology PETRONAS, 32610 Seri Iskandar, Perak, Malaysia

[b] University Malaya Centre of Addiction Science (UMCAS), Faculty of Medicine, Department of Psychological Medicine, University Malaya, Aras 21, Wisma R&D University Malaya, Jalan Pantai Bharu, 59200, Kuala Lumpur

[c] Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 23 September 2015
Revised 25 April 2016
Accepted 27 April 2016
Available online 30 April 2016

Keywords:
Alcohol addiction
Alcohol dependence
Alcohol abuse
Alcohol use disorder
Screening
Logistic model trees
Classification

ABSTRACT

Alcohol use disorder (AUD) has been considered as a social and health issue worldwide. More importantly, the screening of AUD patients has been challenging due to the subjectivity imparted by self-test reports. Automated methods involving neuroimaging modality such as quantitative electroencephalography (QEEG) have shown promising research results. However, the QEEG methods were developed only for alcohol dependents (AD) and healthy controls. Therefore, this study sought to propose a machine learning (ML) method to classify 1) between alcohol abusers and healthy controls, and 2) among healthy controls, alcohol abusers, and alcoholics. The proposed ML method involved QEEG feature extraction, selection of most relevant features, and classification of the study participants into their relevant groups. The study participants such as 12 alcohol abusers (mean age 56.70 ± 15.33 years), 18 alcoholics (mean age 46.80 ± 9.29 years), and 15 healthy controls (mean 42.67 ± 15.90 years) were recruited to acquire EEG data. The data were recorded during 10 minutes of eyes closed (EC) and eyes open (EO) conditions. Furthermore, the EEG data were utilized to extract QEEG features such as absolute power (AP) and relative power (RP). Methods such as *t*-test and principal component analysis (PCA) were employed to select most relevant QEEG features. Finally, the discriminant QEEG features were used as inputs to the classification models: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Multilayer back-Propagation Network (MLP), and Logistic Model Trees (LMT), supported by 10-fold cross validation. As results, the LMT has achieved best performance rendering a classification accuracy (96%), sensitivity (97%) and specificity (93%). In addition, a further classification for each subgroup of AUD patients has achieved accuracy ($> 90\%$). In conclusion, the results implicated significant neurophysiological differences among alcohol abusers, alcoholics, and controls. Moreover, the AUD patients exhibited significantly decreased theta as compared with the healthy controls.

1. Introduction

Alcohol misuse is a common social issue, and has been associated with considerable suffering, mortality, and economic costs [45]. By definition, consumption of alcohol more than 48 g per day or 144 g/week is considered as unsafe [49]. Alcohol intake may be classified as heavy drinking such as alcohol abuse (AA), and a more severe form such as alcohol dependence (AD). According to DSM-IV, both AA and AD are described as a severe form of alcohol abuse causing distress or harm [6]. Furthermore, AA is the recurring use of alcohol despite its negative consequences such as social, interpersonal, and legal problems. On the other hand, the AD or alcoholism is characterized by an increased tolerance and physical dependence on alcohol. In this paper, study participants with AA and AD are referred as *alcohol abusers* and *alcoholics*, respectively. In addition, according to DSM-IV, both AA and AD are commonly termed as alcohol use disorder (AUD).

Conventionally, screening and assessment of alcohol related problems are mainly based on self-test reports [57]. However, the accuracy of self-test reports has been questioned, especially, for heavy drinkers and alcohol abusers [42]. Because the self-test reports may misguide the diagnosis due to patient's memory loss (the patients cannot measure their amount of alcohol consumption) and a dishonest behavior [51,61,70]. In some cultures, discussing alcohol consumption is a social 'taboo'. For example, Solomon [61] and Watson [70] concluded that people tend to

* Corresponding author. Fax: +605 365 7443.
E-mail addresses: wajidmumtaz@gmail.com (W. Mumtaz), aamir_saeed@petronas.com.my (A.S. Malik).

http://dx.doi.org/10.1016/j.knosys.2016.04.026
0950-7051/

under-report alcohol consumption. As a result, objective methods are required to quantify the alcohol consumption and that could also confirm the self-test reports.

Recently, physiological features extracted from electrocardiogram (ECG) data combined with machine learning (ML) tools have been proposed to predict the occurrence of sudden cardiac death [25]. Neuroimaging modalities such as electroencephalography (EEG) could be helpful to objectively quantify the AUD patients. For example, studies involving EEG has shown promise in discriminating alcoholics and control subjects [1,4]. Moreover, the use of visibility entropy to analyze alcoholics EEG data [74] and development of automated techniques to identify alcohol dependence [21,22,27]. More recent research in using EEG for the detection of alcoholism and epilepsy are studied elsewhere [3,35,36]. In addition, the decision support systems are common now and EEG-based methods are not an exception [2]. An increase of theta is studied to reflect a state in which the ability to encode new information is reduced or even blocked [33]. Many QEEG studies are conducted to observe differences of brain activities involving alcoholics and normal subjects [15,16,47,58,72]. For example, Rangaswamy [54] has found an increase of theta in all scalp loci of alcoholics. In contrary, decreased slow band (delta, theta) powers are also reported that could classify alcoholics and controls [58,72,15]. In addition, the decrease of theta power is correlated with the cortical atrophy as seen in MRI of alcoholics [15]. Saletu [58] also claimed that these changes in theta power reflect a hyper-arousal of the central nervous system (CNS). These findings are in-line with previous findings of theta power in alcoholics [59]. The changes in theta power did not seem to be either relate to medication use or presented in the offspring of alcoholics.

Among alcoholics, a decrease in alpha power is associated with deficiency in attention and mental concentration [18]. Ehlers et al. [33,17] reported correlation among EEG voltage and ethnic heritage, family history, gender and other clinical features. According to a study, alcoholics are associated with low voltage ($< 10\,\mu V^2$ per octave) or less power in the alpha frequency range. Besides that, ethnic stratification is possible based on EEG alpha observed from both alcoholics and controls [18] such as EEG low-voltage alpha (LVA) may aid in linking brain function with genetic factors underlying alcoholics. In other words, people with such abnormal EEG LVA will be more prone to alcohol addiction than normal people. Some studies have reported a decrease in alpha relative power (RP) commonly found in alcoholics as compared with controls [58,72,15].

Beta power might be regarded as an electrophysiological indicator of the imbalance in the excitation–inhibition homeostasis in the cortex which was associated with the behavior of alcoholics [65,69]. Studies based on beta relative power elevation in alcoholics, particularly the increase of fast beta power (> 20 Hz) in relapsers, have been well documented [8,55,58,72,15]. However, the increase of beta power is not only the result of chronic and heavy drinking but also found to be related with family history of alcoholic, medication (especially benzodiazepine), sensoperceptual alterations (such as hallucinations, illusions), and clinical seizures [15].

Despite the rapid progress in AUD related studies, very few report on the use of QEEG based machine learning (ML) methods to screen the alcoholic conditions. To our knowledge, some studies have shown high accuracies while utilizing different QEEG features for relapse prediction [72,8], and identification of alcoholics [4]. However, investigations for a full range of AUD patients (alcoholics, alcohol-abusers and a controls group) have been missing. Due to composite nature of the EEG data, it would be helpful to investigate different frequency bands for computation of QEEG features. As they provide useful information regarding the cognitive deficits and altered neural response, which would be helpful to analyze the patient's suffering from Alcohol abuse. Based on these evidences, we have analyzed alpha and theta power as features which are input to the proposed ML method to discriminate the three study groups: AA, alcoholics and healthy controls.

To our knowledge, the proposed ML scheme has never been tested before. This allows automation of the diagnosis process and provides a quick analysis of control and aberrant conditions related to alcohol problems. Due to composite nature of the EEG data, it would be helpful to investigate different frequency bands for computation of features. As they provide useful information regarding the cognitive deficits and altered neural response, which would be helpful to analyze the patient's suffering from Alcohol abuse. In our proposed ML scheme, the feature extraction is followed by feature selection involving PCA and t-test. The classification is based on classifiers such as linear discriminant analysis (LDA), Logistic Model Trees (LMT), Pegosos, and Multilayer Perceptron (MLP), in order to give a performance comparison to check their suitability. Finally, the model validations were provided by cross-validation to avoid classifier over-fitting.

2. Materials and methods

2.1. Study participants

The experiment design was approved by ethics committee, University Malaya Medical Center (UMMC). The data acquisition was performed at UMMC and clinic Bingkor in Kota Kinabalu, Sabah, Malaysia. The study participants included 12 alcohol abusers (mean age 56.70 ± 15.33 years), 18 alcoholics (mean age 46.80 ± 9.29 years), and 15 healthy controls (mean 42.67 ± 15.90 years). The participation was voluntary and each participant had signed consent forms of participation. The experiment design was briefed to the selected participants. Clinical questionnaires such as alcohol use disorder identification test (AUDIT) and MINI international neuropsychiatric interview (MINI), from University Malaya Centre of Addiction Sciences (UMCAS) were used to evaluate the drinking status of the study participants [7,24]. The participants were divided into 2 groups: 1) healthy controls based on alcohol consumption score (first three questions of AUDIT) less than 4 and a total AUDIT score that was less than eight [10], 2) AUDs based on total AUDIT score greater than seven. A further categorization into alcoholics and alcohol abusers were performed based on MINI [39].

2.2. EEG data acquisition

Acquisition of EEG data involved hardware systems from the Discovery 24E and the Enobio system. The Discovery 24E system had 24 channels including: 19 EEG channels (FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2), 1 ground channel, 2 reference channels and 2 for synchronization of events. EEG data were recorded with link-ear (LE) reference. The electrode placement follow the international 10–20 system [32] with amplitude unit in microvolts. Electrode impedance of the electro-gel sensors was maintained below 10 kΩ. The second system, the Enobio was a wireless system with 19 data channels similar locations to the Discovery 24E system, with one external input channel and two electrodes for mastoid references. The EEG data recorded with 2 different systems should have a common reference; therefore, the data were re-referenced to a common reference, i.e., common average reference (CAR).

In BrainMaster system, the data were recorded with sampling rate of 256 samples per second. For Enobio system, the signals were transmitted through Bluetooth protocol at sampling rate of 500 samples per second. The Enobio system used nano-volts as amplitude unit. To make the data compatible between the two EEG