Objective diagnosis of ADHD using IMUs

Niamh O’Mahony a,∗, Blanca Florentino-Liano b, Juan J. Carballo b, Enrique Baca-García b,c, Antonio Artés Rodríguez a

a Department of Signal and Communications Theory, Universidad Carlos III de Madrid, Avda. de la Universidad, 30, Leganés 28911, Spain
b Fundación Jiménez Díaz and Universidad Autónoma (CIBERSAM), Avda. Reyes Católicos, 2, Madrid 28040, Spain
c Department of Psychiatry, New York State Psychiatric Institute and Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 12 May 2013
Received in revised form 19 January 2014
Accepted 20 February 2014

Keywords:
Objective diagnosis
Inertial sensors
Attention deficit/hyperactivity disorder
Machine learning
Classification

ABSTRACT

This work proposes the use of miniature wireless inertial sensors as an objective tool for the diagnosis of ADHD. The sensors, consisting of both accelerometers and gyroscopes to measure linear and rotational movement, respectively, are used to characterize the motion of subjects in the setting of a psychiatric consultancy. A support vector machine is used to classify a group of subjects as either ADHD or non-ADHD and a classification accuracy of greater than 95% has been achieved. Separate analyses of the motion data recorded during various activities throughout the visit to the psychiatric consultancy show that motion recorded during a continuous performance test (a forced concentration task) provides a better classification performance than that recorded during “free time”.

© 2014 IPEM. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common childhood psychiatric disorders and represents a significant public health problem. Several works [1–5] have illustrated a significant increase in the rates of ADHD during recent decades. Thirty years ago, reported prevalence of ADHD among schoolchildren lay between 1% and 3%, whilst recent estimates range between 3 and 10%. Rising trends do not inevitably change in the prevalence of ADHD among youth [5]. These trends in ADHD diagnosis could be due to greater healthcare-seeking behaviours among families with children suffering from ADHD or to higher appreciation of this psychiatric disorder among healthcare providers, parents, and school staff [5]. However, other authors have pointed out that the criteria and rating scales used to diagnose ADHD are, by their nature, subjective and have suggested that difficulties in making an objective diagnosis of the disorder may have contributed to the secular increase in the prevalence of ADHD [6].

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM IV TR) [7], ADHD is characterized by persistent inattention and/or hyperactivity-impulsivity on a more frequent or severe scale than that expected at a particular level of development, and adversely affecting at least two areas of life (e.g. at home and at school) in terms of social, academic or occupational functioning. At present, ADHD is a clinical diagnosis. The American Academy of Child and Adolescent Psychiatry recommend the use of clinical interviews with parent(s) and patient and reports of functioning in school or day care, along with assessment for comorbid psychiatric disorders and review of medical, social and family history of the patient, in the assessment of a child for ADHD [8]. There is no mandatory requirement for psychological or neuropsychological tests [8], despite their usefulness in some cases [9]. New techniques, such as neuroimaging, are becoming a helpful research instrument in the study of ADHD but are not considered useful for the diagnosis [10].

Given that the symptoms of ADHD are non-specific and, usually, are present in some situations but not in others, difficulties in conducting a differential diagnosis exists [11]. In this sense, relying only on DSM IV criteria during the diagnostic process continues to be a subject of heated debate [12]. Not surprisingly, attempting to develop new instruments for an objective diagnosis of ADHD has become a popular research topic, in both the clinical [6,13,14] and biomedical engineering fields [15–17]. Some of the objective tools which have been investigated in the literature to date include continuous performance tests (CPTs) [14,18] and electroencephalogram (EEG) [15]. EEG and CPT testing can be costly to administer, requiring expensive equipment and specialized personnel, especially in the case of EEG.

It has previously been shown that an increased activity level is characteristic of ADHD subjects, in comparison with their
non-ADHD peers [19,20] and, in recent years, sensor technology has evolved to the point where miniature wireless inertial sensors, as small as a wristwatch, can record data for long time periods (up to 24 h). This development has paved the way for the long-term observation of subjects, in order to characterize their habitual behaviour. Recently, these advances have begun to be exploited in the literature, with a number of investigations into the use of accelerometry as an objective tool for diagnosing ADHD [17,21]. In this work, the measurement of the levels and patterns of movement in children, with and without an ADHD diagnosis, is taken one step further: instead of accelerometry, alone, inertial measurement units (IMUs), comprised of both accelerometers and gyroscopes, are used to analyze and characterize the subjects’ motion. The results achieved, in fact, suggest that the inclusion of gyroscope measurements (of angular motion) provides a better discriminative ability between ADHD and non-ADHD subjects than the use of accelerometer measurements (of linear motion) alone.

This proof-of-concept study observes subjects for a duration of approximately 1 h, while they visit a psychiatric consultancy. The subjects are observed both with and without their parents and the psychiatrist, as well as during a CPT task, which requires concentration. Thus, a variety of different situations are observed, representing some of the variety in daily activities, such as school and home. The goal of the study is to determine if inertial measurements, recorded in the setting of the psychiatric consultancy, have the predictive capability to distinguish between ADHD and non-ADHD cases. To this end, machine learning tools are employed and it is shown that good predictive capability is, indeed, attained.

2. Background

2.1. State-of-the-art ADHD diagnosis

The state-of-the-art diagnostic instruments described below were employed in this study to provide the basis for comparison of the proposed method to current best-practice methods.

2.1.1. ADHD Rating Scale-IV: Home Version (ADHD RS-IV) [22]

The ADHD RS-IV consists of eighteen items that assess DSM-IV criteria for inattention and hyperactivity. Good validity, test-retest reliability and internal consistency have been demonstrated for the rating scale [23].

2.1.2. Test of variables of attention (T.O.V.A.®) [18]

The T.O.V.A.® is a CPT, consisting of a 21.6 min long test, during which subjects have to respond quickly to targets whilst withholding response to non-targets. It should be noted that the T.O.V.A. is not intended as a stand-alone factor for diagnosis. Indeed, the accuracy of the T.O.V.A.® alone is not sufficient for reliably distinguishing between ADHD and non-ADHD subjects, with the T.O.V.A.® Clinical Manual reporting sensitivity and specificity of 84% and 89%, respectively, using a discriminant analysis method. Independent studies have shown similar or worse results, e.g. [24]. This performance is not, however, specific to the T.O.V.A.® and other available CPTs have a similar performance [6,25].

2.1.3. Clinical diagnosis by psychiatrist

Clinical evaluation of ADHD and other psychiatric disorders included reviewing the clinical interview and all available data, following the recommendations of the AACAP, as detailed in Section 1 [8].

In order to estimate the predictive sensitivity and specificity of any test for ADHD, the outcomes have to be compared with the outcomes of the same patients using alternative tests. Given the lack of a non-subjective and universally agreed-upon “Gold standard” test, it is often unclear what is considered as the ground truth for a correct ADHD diagnosis. In this work, a combination of the ADHD RS-IV rating and the clinical diagnosis serves as the ground truth diagnosis for the study, such that only those patients with a positive ADHD diagnosis in the ADHD RS-IV scale and for whom this positive diagnosis was confirmed according to the psychiatrist’s evaluation are considered as ADHD subjects, whilst those with a negative diagnosis in the ADHD RS-IV scale, confirmed by a negative clinical diagnosis are considered as controls. Any subject with contradictory diagnoses was eliminated from the study. Furthermore, patients who are undergoing treatment for their ADHD (i.e. pharmaceutical treatment) were not considered, as the level of effectiveness of their medication can vary and this could result in a falsely heterogeneous ADHD group.

2.2. Inertial measurements and activity

Recent advances in sensor technology have lead to the widespread availability of affordable, miniature inertial sensors which can be comfortably worn by a human subject, going about normal daily activities. Currently available miniature inertial sensors typically consist of tri-axial accelerometers and tri-axial gyroscopes to measure the total inertial force and angular velocity, respectively, on mutually perpendicular axes (x, y and z). The total inertial force provides a representation of the subject’s pose (inclination angle) and linear motion. The angular velocity provides a measure of the rotational characteristic of the subject’s motion. As such, these sensors allow both a quantitative and qualitative characterization of the subject’s motion, measuring both the level or intensity of movement and the nature (rotational, linear). This characterization of motion is exploited in this work to identify differences in movement-related behaviour between subjects who suffer from ADHD and those who do not.

For more details regarding the tri-axial accelerometers and tri-axial gyroscopes used in this work, the reader is referred to the APDM Opal sensor product information contained in [26]. These sensor modules save data directly to an on-board memory for post-processing.

2.3. Machine learning

2.3.1. Classification

In this work, the determination of the predictive capability of inertial measurements in ADHD diagnosis, is based on machine learning methods, specifically, classification methods. Many classification algorithms exist in the literature; the one used in this work is the support vector machine (SVM), a state-of-the-art learning machine, used in a wide variety of applications. For the sake of brevity, the theory of operation of the SVM will not be described here – instead the reader is referred to [27] for details. In this work, the SVM is simply used as a tool to evaluate the classification performance of features calculated from the inertial measurements.

2.3.2. Feature selection

Feature selection methods are used to reduce dimensionality of data for various reasons: to reduce computational load, to improve the generalization ability of the classifier and to improve the interpretability of the results. A good review of feature selection techniques can be found in [28]. In this work, a suboptimal forward selection method is employed as a tool to identify those features which have the best predictive capability.
3. Methods

3.1. Subjects

Forty-three children, aged 6–11 years, who were referred to the Child and Adolescent Psychiatry Unit of the Department of Psychiatry at Fundación Jiménez Díaz Hospital (Madrid, Spain), were included in this study. The experimental group consisted of children diagnosed with ADHD (N = 24, 55.8%), and the control group consisted of children not diagnosed with ADHD (N = 19, 44.2%). All subjects in the experimental group, but none in the control group, met DSM-IV criteria for ADHD. In the experimental group, 45.8% met criteria for the ADHD combined subtype, 41.7% for the predominantly inattentive subtype, and 12.5% for the hyperactive/impulsive subtype. In the experimental group 70.8% were male, whilst 47.4% were male in the control group. Mean age in the control group was 9.05 (±1.39), whereas mean age in the experimental group was 8.54 (±1.38) (t-test: df = 41, p = 0.24).

3.2. Ethics procedures

Parents and children were given a detailed description of the project and each subject’s parent(s) or guardian(s) were required to give signed informed consent. Subjects provided assent. The consent and assent forms and the study protocol were reviewed and approved by the Institutional Review Board of Fundación Jiménez Díaz Hospital.

3.3. Monitoring procedure

The forty-three subjects were all evaluated in the same fashion. Two IMUs were used to record movement data for each subject – one attached to a belt worn at the waist and the other fixed by a velcro strap to the ankle of the “dominant” leg (the subject was asked to kick a football to identify the dominant leg). The IMU at the waist measures movement of the whole body, whilst the IMU at the foot captures local movements such as tapping the foot or other behaviour which may be associated with excess energy. Each subject wore both IMUs during the entire duration of their visit to the psychiatric consultancy (approximately 1 h). A trained employee (the “supervisor”) accompanied each subject throughout the entire visit, labelling on a time-synchronized computer platform the “context” in which the data were collected, where the context referred to the “where, what and with whom” of the subject’s environment. The subjects spent time in each of the following contexts:

- **WP**: Waiting room; with parents.
- **WS**: Waiting room; with supervisor only.
- **CD**: Consultant’s room; with psychiatrist.
- **CP**: Consultant’s room; with psychiatrist and parents.
- **TT**: Taking the TOVA test; with supervisor only.

In both the waiting room and the consultant’s room, the subjects were requested to sit in a revolving chair. While taking the TOVA test, the subjects were seated in a non-rotating chair in a room containing only a desk with the computer for administering the T.O.V.A. and nothing on the walls, to eliminate distractions.

A minimum duration of 2 min in each context was selected as valid, for the purpose of data analysis, to reduce the effects of interruptions and disruptions and to allow a fair measurement of the subjects’ behaviour in each context.

3.4. Feature calculation and selection

In order to represent the recorded data in a concise form, features were calculated from the observed accelerometer and gyroscope signals. In order to exploit the most representative feature set possible, characterizing all aspects of the observed signals, features were calculated per sensor type (i.e. for gyroscope and for accelerometer), for each sensor location (waist and foot) and during each context. Features were calculated based on the modulus of the tri-axial accelerometer and gyroscope vectors, as well as on differential data (e.g. the difference between the modulus of the acceleration vector for consecutive samples). Additionally, age, gender and the achieved T.O.V.A. score were treated as additional features for each subject. The entire list of features is not included, for the sake of brevity, but the main categories, with some examples and the number of features in each category, are listed in Table 1.

A forward selection method was employed to select the subset of up to a maximum of 15 features which achieved the best classification accuracy for each context. For each context, the subset which achieved the maximum performance with the minimum number of features was selected as the final feature subset for that context. Leave-one-out validation was employed in the training and test phases of classification, due to the small number of observations available. The same forward selection procedure was also repeated, considering the set of all features calculated throughout the entire duration of the experiment (i.e. 668 features per subject).

### Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>High resolution histograms [21]</td>
<td>35</td>
</tr>
<tr>
<td>Correlation between sensors/sensor types</td>
<td>8</td>
</tr>
<tr>
<td>Basic statistics: (e.g. mean, std. dev., skewness, kurtosis)</td>
<td>52</td>
</tr>
<tr>
<td>Frequency domain: (e.g. low/high frequency power ratio.)</td>
<td>8</td>
</tr>
<tr>
<td>Nonlinear features: [17] (e.g. central tendency measure)</td>
<td>4</td>
</tr>
<tr>
<td>Structural features: [15] (e.g. positive/negative area, slope, peak-peak value)</td>
<td>20</td>
</tr>
<tr>
<td>Motion features: [20] (e.g. number of movements per second)</td>
<td>6</td>
</tr>
<tr>
<td>Total per context</td>
<td>133</td>
</tr>
<tr>
<td>T.O.V.A. score, age, gender</td>
<td>3</td>
</tr>
<tr>
<td>Total per subject</td>
<td>668</td>
</tr>
</tbody>
</table>

4. Results

Two different SVMs were used to classify the subjects - one with a linear kernel and the other with a Gaussian kernel. However, despite an improved classification accuracy being achieved by the Gaussian kernel SVM, results for the linear SVM only are reported in this section, because, due to the small number of samples available in the dataset, the Gaussian SVM may be prone to overfitting the training data and giving an unrealistically elevated estimate of the classification performance.

Table 2 shows the classification performance achieved for the set of k features which resulted in the highest classification accuracy in the forward selection procedure. Results are shown for each context of the trial individually and for the entire duration of the trial (“ALL”). The following performance measures are reported: accuracy (percentage of correctly classified subjects), specificity (percentage of correctly classified non-ADHD subjects) and sensitivity (percentage of correctly classified ADHD subjects).

### Table 2

<table>
<thead>
<tr>
<th>Context</th>
<th>k</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP</td>
<td>5</td>
<td>83.72</td>
<td>73.68</td>
<td>91.67</td>
</tr>
<tr>
<td>WS</td>
<td>14</td>
<td>85.37</td>
<td>89.47</td>
<td>81.81</td>
</tr>
<tr>
<td>CD</td>
<td>7</td>
<td>83.72</td>
<td>73.68</td>
<td>91.67</td>
</tr>
<tr>
<td>CP</td>
<td>5</td>
<td>81.40</td>
<td>57.89</td>
<td>100</td>
</tr>
<tr>
<td>TT</td>
<td>6</td>
<td>93.02</td>
<td>89.47</td>
<td>95.83</td>
</tr>
<tr>
<td>ALL</td>
<td>10</td>
<td>95.12</td>
<td>94.44</td>
<td>95.65</td>
</tr>
</tbody>
</table>
The performance of the SVM is seen to be good, in terms of accuracy, sensitivity and specificity, across all of the individual contexts and, particularly, during the T.O.V.A. test. The results across the entire duration (ALL) show that an increased classification performance can be achieved by the joint consideration of the subject’s behaviour in multiple contexts.

For the sake of brevity, the entire list of features selected by the algorithm for each of the contexts listed in Table 2 is not included here. In summary:

- Approximately one-third of the features chosen by the feature selection procedure are gyroscope-based and the remainder are predominantly accelerometer-based (in certain contexts, gender, age and T.O.V.A. score have also been chosen).
- More than half of the selected features came from the sensor at the child’s waist.
- Of the ten features chosen by the forward selection procedure across the entire test duration, six of those came from the T.O.V.A. test context.
- More than half of the features chosen in every context were “high resolution histogram” (HRH) elements.
- The other categories of features which were selected most frequently were basic statistics, frequency domain characteristics and measures of correlation between sensors.

5. Discussion

The results, in Section 4, suggest that inertial sensors provide a promising tool for the objective diagnosis of ADHD. In the previous literature, the focus has been on acceleration measurements. Interestingly, an analysis of the features selected for each context suggests that the features calculated based on gyroscope measurements have a good predictive capability for discrimination between ADHD and non-ADHD subjects and should be considered along with accelerometer-based features. More than half of the selected features in all contexts came from the sensor at the child’s waist, suggesting that “global” motion of the body is a better indicator of hyperactivity than “local” motions, such as tapping the foot, or similar habits.

Interestingly, of the ten features chosen by the forward selection procedure across the entire test duration, six of those came from the T.O.V.A. test context. Thus, it appears that the movement of a child in restrictive situations, such as carrying out a specifically defined task that requires concentration (the T.O.V.A.), provides better discriminative ability than the child’s behaviour in “free time”. Furthermore, it should be noted that the sensitivity and specificity of the reported method outperform the same metrics for the T.O.V.A. alone, which were 84% and 89%, respectively, as cited in Section 2.1.2.

The final observation, regarding the chosen features at the output of the forward selection, pertains to the categories to which chosen features belong. More than half of the chosen features in every context were “high resolution histogram” (HRH) elements. This does not come as a surprise, since a recent review of the state-of-the-art in extraction of features from accelerometers for the ADHD problem [17] showed that HRH features provide the best classification performance among all of the reviewed categories. However, it should be noted here that other categories of features have also been chosen in all contexts, most notable among them, basic statistics, frequency domain characteristics and measures of correlation between sensors. Thus, it can be seen that using a wide range of feature types provides improved performance over methods which rely on one single category of features.

6. Conclusions and future work

This work serves as a proof-of-concept for the use of accelerometers and gyroscopes in the diagnosis of ADHD in children between 6 and 11 years of age. Results have shown that classifiers based on a small number of features (5–10) can discriminate between ADHD subjects and control subjects with high sensitivity and specificity levels. Clearly, a more extensive follow-up study is required, with a larger number of patients, to more robustly identify the optimal set of features and the diagnostic accuracy of the system. To this end, construction of a large database has already been begun by the authors.

Furthermore, investigation into the ongoing monitoring of patients with an existing ADHD diagnosis, based on the inertial sensors and classification techniques described in this work, is also underway. The aim of that study is to use objective measures to provide feedback on the progress of the patient, their response to treatment and other indications.

Funding

This work has been partly supported by Ministerio de Educación de Spain (projects ‘DEIPRO’, id. TEC2009–14504-C02-01, and ‘COMONSENS’, id. CSD2008-00010) and partly supported by an Educational Grant to the Department of Psychiatry, Fundación Jiménez Díaz Hospital.

Ethical approval

Ethical Approval was given by the Comité Ético de Investigaciones Clínicas (equivalent to IRB). Judgements reference number is: PIC51/2011.

Conflict of interest

None declared.

References


