Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity

H.J. Kam[1]; Y.M. Shin[2]; S.M. Cho[2]; S.Y. Kim[2]; K.W. Kim[3]; R.W. Park[1]

[1]Department of Biomedical Informatics, Ajou University School of Medicine, Wonchon-dong, Yeongtong-gu, Suwon, South Korea;
[2]Department of Psychiatry and Behavioral Science, Ajou University School of Medicine, Wonchon-dong, Yeongtong-gu, Suwon, South Korea; [3]Neuropsychiatry and Stroke Center, Seoul National University Bundang Hospital, Seongnam, South Korea

Summary

Objective: Questionnaire-based ADHD screening tests may not always be objective or accurate, owing to both subjectivity and prejudice. Despite attempts to develop objective measures to characterize ADHD, no widely applicable index currently exists. The principal aim of this study was to develop a decision support model for ADHD screening by monitoring children's school activities using a 3-axial actigraph.

Methods: Actigraphs were placed on the non-dominant wrists of 153 children for 3 hours, while they were at school. Children who scored high on the questionnaires were clinically examined by child psychiatrists, who then confirmed ADHD. Mean, variance, and ratios of low-level (0.5-1.0G) and high-level (1.6-3.2G) activity were extracted as activity features from 142 children (10 ADHD, 132 non-ADHD). Two decision-tree models were constructed using the C5.0 algorithm: [A] from whole hours (class + playtime) and [B] during classes. Accuracy, sensitivity, and specificity were evaluated. PPV, NPV, likelihood ratio, and AUC were also calculated for evaluation.

Results: [Model A] One child without ADHD was misclassified, resulting in an accuracy score of 99.30%. Sensitivity and NPV were 1.0000. Specificity and PPV were 0.992 and 0.803-0.909, respectively. [Model B] Two children without ADHD were misclassified, resulting in an accuracy score of 98.59%. Specificity and PPV were scored at 0.985 and 0.671-0.832, respectively.

Conclusion: The selected features were consistent with the findings of previous studies. Objective screening of latent patients with ADHD can be accomplished with a simple watch-like sensor, which is worn for just a few hours while the child attends school. The model proposed herein can be applied to a great many children without heavy cost in time and manpower cost, and would generate valuable results from a public health perspective.

Correspondence to:

Rae Woong Park, M.D., Ph.D.
Department of Biomedical Informatics
Ajou University School of Medicine
Wonchon-dong, Yeongtong-gu, Suwon,
Gyeonggi-do, 442-749, South Korea
Phone: +82-31-219-5342
Fax: +82-31-219-4472
E-mail: veritas@ajou.ac.kr

1. Introduction

Attention-deficit hyperactivity disorder (ADHD) is the most common neurobehavioral disorder of childhood, and is among the most prevalent of mental health problems, having been estimated to affect 3-9% of school-aged children [1-3]. ADHD is characterized by difficulties with attention, impulse control, and hyperactivity relative to typical children of the same age and gender. ADHD has been implicated in up to 10-fold increases in the incidence of antisocial personality disorder, and up to 5-fold increases in the risk of drug abuse [1]. Therefore, early diagnosis and treatment are necessary for children who suffer from ADHD. Regardless of the diagnostic skill of the clinician, the signs of ADHD are notoriously variable and situational in nature [4]. A single assessment over a brief visit may not be reflective of the patient's general level of activity or impairment [5]. Consequently, clinicians frequently rely heavily on impressions and scorings from schoolteachers or parents, which are neither consistently objective nor accurate [1].

Therefore, in this study, actigraphs were applied in an effort to non-invasive observe long-term changes in children's activity [6]. The primary advantage in using an actigraph is that it allows for the acquisition of a patient's information in natural settings. Activity is extremely complex, and thus activity data reflects continuous and multi-dimensional positional changes. An actigraph is an electronic device that simplifies and quantifies this complex information into numerical values, which are studied extensively in the artificial intelligence and pervasive computing communities [7-10]. Actigraphs have also been used in several studies to evaluate alterations in body position [11-13]. Posture and motion patterns, according to many researchers and theorists, are key indicators of emotional state [14]. An actigraph, then, can also be used to record motor activity in individuals suffering from psychiatric disorders; this approach has already been applied to the studies of individuals with Parkinson's disease and akathisia [15-17].

Children suffering from ADHD tend to be more active, restless, and fidgety than typical children. A number of previous studies have been conducted to develop objective measures, define hyperactivity, and quantify disturbances in attention or impulse, as compared to controls, by employing the behavioral characteristics evidenced by patients with ADHD. Activity scores were compared with the scores of continuous performance tests with activity [18]: The results demonstrated significant differences in activity levels between ADHD/non-ADHD groups. Many actigraphy studies have corroborated this significant difference in activity between children with/without ADHD. Children with ADHD have been shown to be 20-30% more active than normal children in school- or laboratory-based attention tasks, as determined by analyzing data generated by wearable actigraphs [19-20]. Tsujii *et al.* attempted to determine whether any association could be drawn between activity level and situational factors [21]. In their study, patients with ADHD and controls wore actigraphs for 1 week while attending school, and the maximum difference in activity level between the two groups was noted when the effects of inhibition and fatigue overlapped ADHD.

Preexisting accelerometers were used to actigraphically assess subjects' activities over the long term, by generating counts or summaries of activity in excess of a pre-defined temporal threshold. These long-term measurements rendered it possible to evaluate subjects' activity levels quantitatively. Recently, Wood *et al.* [22] reported that ADHD/non-ADHD subjects differed significantly in terms of the number and the intensity of their movements. The subject children's activities were measured via three measurements: the number of movements, the magnitude of those movements, and the intra-individual variability by calculating standard deviations for each minute. The methodology applied by Wood *et al.* is the most similar to the new high-resolution activity features we described herein; however, this approach is somewhat limited in that it cannot adequately address multidimensional (or high-resolution) changes of activity in a given epoch, owing to the compressions or reductions of dimensionality that would be required for such measurements. Moreover, earlier studies were generally primarily focused on statistical differences between ADHD/non-ADHD groups, and thus there were no indications regarding the extent or quality of differences in figures by which one could derive a distinction between ADHD/non-ADHD groups. Numerous attempts have been made to develop objective measures for a definitive characterization of ADHD. However, these efforts have not thus far resulted in a widely applicable standardized index.

© Schattauer 2010

H. J. Kam et al.: Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity

2. Objectives

The principal objective of this study was to develop a decision support model for ADHD screening by monitoring children's activities at school using a 3-axial accelerator (actigraph).

3. Methods

3.1 Participants and Clinical Assessment

A total of 153 children (78 boys and 75 girls; range 7-9 years) were recruited from a regular elementary school. Questionnaires using the K-CBCL (Korean Child Behavior Checklist) and K-ARS (Korean ADHD Rating Scale-IV) were administered to the subject children's parents and homeroom teachers. The CBCL is an extensively utilized instrument for evaluating various aspects of a child's behavior, as observed by the child's parents. The CBCL consists of a 113-item parent-report questionnaire which surveys the internalizing/externalizing symptomatology of a child. The ARS was designed by DuPaul *et al.* [23] to evaluate ADHD symptom severity according to the DSM-IV, using a 4-point rating scale ranging from 0 to 3. In general, the top 10% of scores are selected as the high-risk group. In this study, children who scored more than 60 points on the K-CBCL questions or scored in the upper 10% on the K-ARS questions were selected as the high-risk group.

Children in the high-risk group were subsequently interviewed and clinically diagnosed via close examination by four experienced child and adolescent psychiatrists. 11 children who refused to be clinically examined were excluded from further study. The interviews included K-SADS-PL-K (Kiddie Schedule for Affective Disorder and Schizophrenia-Present and Lifetime Version-Korean Version) based on the DSM-IV and mentality test. The K-SADS-PL is a semi-structured interview tool designed to evaluate the severity of ADHD symptoms and the current and affected state throughout the lifetime of the child for 32 different psychiatric disorders included in the DSM-IV, using objective diagnostic criteria. The Korean version of the K-SADS-PL was translated by Kim *et al.* [24], and its validity and reliability have been well established for the assessment of ADHD, tic disorders, and oppositional defiant disorder. The clinical diagnoses of 24 children in the high-risk group were conducted in two steps. A primary interview and diagnosis was carried out by two experienced child and adolescent psychiatrists who were well-trained in the use of K-SADS-PL-K; and a final clinical diagnosis was conducted by two senior child and adolescent psychiatrists through discussion. Children diagnosed with ADHD were regarded as gold standards for this study. The study was conducted as a double-blind model; no one involved in the acquisition of activity data was informed in any way of the results until the end of the experiment (Part A at ▶Figure 1).

3.2 Elementary School in Korea

In Korea, children between the ages of 6 to 12 attend mandatory elementary schooling. The Korean elementary curriculum consists of 40-minute courses followed by 10-minute recess periods, with the standard courses of study including native language, math, science, ethics, art and physical education (as well as some additional grade-specific subjects). For the first- and second-grade students who were the subjects of this study, the first class begins at 9 a.m. and the final (fourth) class ends at 12 p.m. Generally speaking, children are assigned seats within a classroom, and thus there is no need to change seats or classrooms except under special circumstances. Approximately 30-40 children attend each class with the same teaching materials, while remaining seated in their own assigned seats. One homeroom teacher is assigned to each class, and that teacher maintains sole responsibility for the class.

3.3 Activity Measurement

Activity data was collected for 1-3 days, for 3 hours per day, during school hours. An actigraph (LIG Nex1 Co., Ltd., South Korea, 2007) was placed on the non-dominant wrist of each child from the beginning of the first lesson to the end of the final lesson. Opinions vary regarding the 'proper'

sensor attachment sites for the acquisition and analysis of children's activity [4, 6, 18, 19, 21, 25]. Children are generally assigned seats within a single classroom, and thus seat or classroom changes are rarely necessary except under special circumstances. The wrists appeared to be a more generally appropriate attachment site than the torso (waist or chest) or ankles, considering the aspect of reflection for various types of activity. Additionally, the wrist-worn sensor type also has the advantage of easier management when attempting to measure large groups of children (four classrooms with approximately 40 students per class). 3-axial acceleration data were recorded in a built-in memory chip at a frequency of 32 Hz.

3.4 Feature Extraction and Selection

From the data acquired over the 1-3 day(s) collection period, data from the first valid day was taken for each student. Vector-summed absolute activity values (unit: gravity [G]) were calculated from the 3-axial activity data. Afterward, the following overall and 1-minute epoch features were extracted: mean, variance, and ratios of low-level (0.5-1.0 G) and high-level (1.6-3.2 G) activity over the entire activity period, as shown in ▶Figure 2. Features of high-level activity increased at 0.2 G intervals from 1.6 G, and features of low-level activity decreased at 0.1 G, beginning at 1.0 G. Here, a ratio for a certain activity level means activity counts in a certain activity acceleration region (in gravity) over the whole activity counts, with the intent to validate the portion of the certain level of activity within the time interval:

$$\text{Ratio of certain activity level}(x) = \sum_{1\,\text{min}} \frac{\text{activity count}(x)}{\text{Counts of whole activity records}} \quad (1)$$

$$\text{Activity count}(x) = \begin{cases} \text{Counts of activity lower than } (x) & \text{if } 0.5 \leq x \leq 1.0 \\ \text{Counts of activity higher than } (x) & \text{if } 1.6 \leq x \leq 3.2. \end{cases}$$

The features were assigned to one of two categories: (1) features from the whole three hours (class + playtime) and (2) features during classes. The term 'during classes' refers specifically to the middle 14 minutes of each class. Additionally, the subjects' gender and grade were added as features.

From the extracted features, we selected a small set of features that were valued highly in terms of information richness for the classification of patients with ADHD. The number of selected features was dependent on the number of samples [26] – in our case, the number of children reflected in a model construction. The WEKA (WEKA 3.5, University of Waikato Hamilton, New Zealand) provides several feature selection options: we utilized information gain methodology for selection features by the 10-fold cross-validated information gain method after the feature extraction process.

3.5 Model Construction for ADHD Screening

We have employed a decision-tree in this study, owing to its ready intelligibility relative to other data mining methodologies for clinicians. Models involving other methods are generally more difficult to understand owing to their black-box decision-making processes. A decision-tree is a supervised classification method exploited in data mining. First, a decision-tree is constructed using a 'training' set of subjects. Then, the constructed model can be utilized as a predictive decision-support model for new subjects. A decision-tree has a flowchart-like upside-down tree structure [27], in which dependent variables -whether or not a subject is in the high-risk group of ADHD – are positioned at the lowest ends of the tree. Nodes are formed when the branches stretch down, where selected features and conditions for the features are placed: they represent the best capability to separate the subjects gathered at a node into their own dependent variable. After model construction, a decision-tree lists several features and conditions by which subjects can be divided into proper dependent variables, and it shows the conditions schematically.

Theoretically, exponentially many decision-trees can be formed by any combination of a given set of attributes (features). Finding the optimal decision-tree is extremely difficult and has no known solution (a 'NP-complete problem'); thus, many algorithms employ heuristic search methods in hypothetical spaces [28]: many statistical algorithms and a variety of commercial software are

© Schattauer 2010

H. J. Kam et al.: Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity

currently available [29]. Matthew *et al.* [29] previously conducted a comparison study on variety of decision-tree algorithms designed to solve problems associated with practical classifications. Among them, the most frequently used algorithms – IDE3 (Iterative Dichotomiser 3), C.4.5/C5.0, CART (classification and regression trees) and SPRINT – were possible candidates for the most appropriate algorithm on ADHD screening. They compared the classification accuracies of the four algorithms with a variety of group sizes (from 2 to 13), numbers of attributes (from 9 to 24) and numbers of records (from 270 to 43,499): the SPRINT showed the best classification accuracy among all of the classifiers, and C4.5 evidenced the next-best classification accuracy, with difference of 0.5%-1.63%. C4.5 has a reported accuracy of 76.7% with 270 records among two groups. An excessively large sample size would results in superior performance; however, this is not an essential prerequisite for the use of using a decision-tree analysis. In the cases of IDE3 and CART, the resulting accuracies were considerably lower: additionally, the variance of the accuracies was relatively high, and was dependent on group size, attribute number and record number. On the contrary, in SPRINT and C4.5, the conditions did not affect the classification accuracy [29]. SPRINT is specifically designed for massive data handling, and provides an extendable parallel algorithm. However, that function was not an essential prerequisite for ADHD screening; in this study, the C5.0, which is the commercial successor of a well-known classifier C4.5, was employed.

In this study, we constructed two tree-shaped screening models using the C5.0 algorithm (Clementine 10.1, SPSS Inc.) according to two groups of features: (1) from whole hours (class+playtime) – Model A and (2) during classes – Model B. The C5.0 is known to offer a number of improvements over C4.5 [30]. From the perspective of resulting performance, C5.0 yields results similar to those of C4.5, but with considerably smaller decision-trees: to avoid errors due to by model over-fitting, simpler models are preferred: this is referred to as Occam's razor or the principal of parsimony [28]. Additionally, the class (group) imbalance problem of the subject groups was resolved by the un-equal misclassification costs. The C5.0 provides the newly-adapted weighting function of variable misclassification cost, which is yet another reason for selecting the C5.0. C5.0 constructs decision-trees based on the concept of information entropy, and assesses the normalized information gain (difference in entropy) that results from the selection of a valuable feature for the separation of the subjects into ADHD and non-ADHD groups. The feature with the highest normalized information gain is the one employed in the decision regarding the downward-stretching branches. The algorithm subsequently recurs on the smaller subset of subjects, thereby resulting in a tree-shaped model [31].

As is the case in many other clinical environments, there exists a group imbalance problem: the non-ADHD group was represented by a large number of examples, whereas the ADHD group was represented by only a few examples. Among many approaches to addressing the group imbalance issue, we applied the misclassification cost, by altering the relative costs of misclassification of the small and the large classes [32]. In constructing the models, we utilized 10-fold cross-validation methodology that made it possible to utilize all observations as training and validation, and each of these observations was used for validation exactly once [31]. (Part B, ▶Figure 1)

3.6 Model evaluation and statistical analysis

To evaluate the performance of the models, we calculated several statistical values with the resultant confusion matrices. Accuracy, sensitivity, and specificity were evaluated for each model. Positive predictive value (PPV), negative predictive value (NPV), likelihood ratio (±), and area under ROC curve (AUC) were also calculated in order to assess performance as a screening model. In order to compare the selected features as tree nodes between groups, Mann-Whitney's U-test was applied for evaluation.

4. Results

4.1 Clinical Assessment

The whole subjects' demographic and clinical characteristics are provided in ▶Table 1. Among the 35 high-scoring (high-risk) students, 24 children were clinically diagnosed, and the confirmed diagnoses were as follows: 10 children (8 boys and 2 girls) were diagnosed as ADHD, 5 children (4 boys and 1 girl) had other psychiatric problems, and the others (3 boys and 6 girls) were non-ADHD. Demographic features of the 24 diagnosed children and their parents are provided in ▶Table 2. Although approximately 33.3% of children with ADHD evidence comorbidity and symptomatic overlap with other psychological problems in the Korean community samples [33], none of the children with ADHD in our samples (n = 10) evidenced comorbidity and symptomatic overlaps: none of them had been previously diagnosed with ADHD or any other psychiatric problems, nor had any taken medications for the treatment of such issues. To develop a screening model for the screening of patients with ADHD, the low-risk group from the screening examination was classified as 'non-ADHD' (n = 132), resulting in a prevalence of 7.04%.

4.2 Selected Activity Features and Constructed Screening Models

By combining the time interval, statistical values, and extents of activities, 106 activity features were extracted from a single day of actigraphic data. We reduced the feature dimensionality such that 20 selected features were valued highly in terms of information richness for classification with WEKA. The selected features, with the exception of gender in Model B, were also made on the basis of 1-min epoch intervals: variance of 1-min mean, variance of 1-min variance, and high-level (2.2-3.2 G) activity ratios were top ranked.

The applied misclassification cost – a weighting method for imbalanced data – for a patient with ADHD (as a 'non-ADHD' child) was found to be 13.2 times that of the misclassification cost of a non-ADHD child. The cost was applied in accordance with the imbalance of ADHD/non-ADHD groups (10:132), and reflected the relative importance of ADHD screening, considering the social costs to be shouldered when proper treatment opportunities are lost. Inverse cost can also become relevant in cases in which children without ADHD are misdiagnosed with ADHD. However, the results of classification are appropriate for screening purposes, rather than for diagnostic purposes. The clinical diagnosis/confirmation of ADHD and the appropriate treatments were carried out by experienced child psychiatrists, independent of the actigraphic data. Two decision-trees were constructed after 10-fold cross-validation, and the structures of these trees are shown in ▶Figure 3.

4.3 Comparing Selected activity features in ADHD and Non-ADHD groups

All of the selected features, with the exception of VLR1.0 (variance of low-level activity ratio [<1.0 G]), were statistically significant at a 95% confidence level (▶Table 3). In all the variants that could be compared, the ADHD group exhibited a higher activity variance or mean than the non-ADHD group. Model A showed a significant difference (p<0.001) in the Vmean (variance of mean) among ADHD/non-ADHD individuals. Other three values, VHR1.8 (variance of high-level activity ratio [>1.8 G]), VHR3.2, and MHR3.2 (mean of high-level activity ratio [>3.2 G]) also differed significantly between groups (p<0.001). In Model B, VLR0.6, VLR0.7, and VLR0.8 showed significant differences (with the exception of VLR1.0). In terms of features of high-level activity, MHR1.6 and MHR2.0 differed more significantly (p = 0.006 and 0.003, respectively).

4.4 Model Evaluation

PPV and NPV changed depending on the prevalence of the disease; therefore, prevalence of 3% and 7% were applied to calculation (▶Table 4). In Model A, one child without ADHD among 142 samples was misclassified. The accuracy was 99.30%, as is shown in ▶Table 4. Sensitivity and NPV were 1.000. Although NPV changed depending on the prevalence of the disease, the patients with ADHD were unequivocally classified into the ADHD category, and the NPV value was 1.000 regardless of

© Schattauer 2010

H. J. Kam et al.: Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity

the prevalence of ADHD. The specificity and PPV values were 0.992 and 0.803-0.909, respectively. Additionally, several other values such as the likelihood ratios and AUC were utilized to estimate the performance of a screening model, and are shown in ▶Table 4. In Model B, the decision-tree with features of class hours yielded inferior results as compared to Model A. Among a total of 142 children, 2 children without ADHD were misclassified; this caused deterioration in quality, and resulted in an accuracy of 98.59%. Specificity and PPV declined slightly, with scores of 0.985 and 0.671-0.832, respectively.

5. Discussion

5.1 Screening Models for ADHD by Monitoring Children's Activity

We hypothesized ADHD evidence distinctive activity properties different from non-ADHD: that makes it possible to formulate a quantitative and objective screening model capable of effectively distinguishing ADHD and non-ADHD. It also proved possible to make some limited distinctions between ADHD and other psychiatric disorders. This study was achieved by a massive data gathering effort, using a specially designed high-frequency actigraph. As is shown in the Results section, the children with ADHD could be separated from the children without ADHD by certain activity features. These features were extracted from children's daily activities in the form of 3-axial acceleration data while the children wore actigraphs for 3 hours; that is, distinctive features in activity were observed that could differentiate ADHD and non-ADHD. Accordingly, we were able to use these features to construct an effective ADHD screening model.

As a model for ADHD screening, we treated the cost of false-negative misclassification as much higher than that of false-positive misclassification. False-negative misclassification, in which a patient with ADHD is incorrectly judged as a 'non-ADHD' child, deprives patients with ADHD of their opportunity for consultation and treatment. As mentioned previously, there are also social costs incurred by false-positive misclassifications; however, the proposed screening model provides primary screening functions, by definition, and is not designed for diagnosis or confirmation. It can be employed as a screening methodology with little repulsion – still, some negative emotions toward visiting psychiatry remain, which can possibly result in delays or loses of treatment opportunities – in daily school life. Therefore, from the healthcare perspective, a false-negative result can cost a great deal more than a false-positive result. As a screening model, sensitivity and NPV can be set at a higher value on evaluation than accuracy. ▶Table 4 shows that the tree models had the most complete sensitivity and NPV scores available, as well as a high accuracy score. Furthermore, specificity and PPV were also not particularly far behind, even though we regarded the cost of false-negative misclassification as much higher than the cost of false-positive misclassification. Value such as the likelihood ratio (+) represent model performance, and the AUC also exhibits a broad capability for ADHD screening.

Additionally, as compared with currently existing questionnaires, which generate subjective indices that can be changed or distorted as the result of subjectivity or prejudice, the tree models proposed herein are objective. The constructed decision-trees can be readily transformed to decision-support rules. That makes it possible to implement a simple device or software for ADHD screening.

5.2 Significance of Selected Features

Using several valid features, we constructed effective decision-trees. In detail, the variance of 1-min means and the mean and variance of high activity (1.8 G and 3.2 G) ratios were the principal factors for group separation in Model A. The other tree with class features evidenced slightly different aspects: gender, variance of low-level activities (0.6-1.0 G) and high-level activity ratios close to the mean activity were applied that were not important in the other model, and the features or statistical values obtained from the 1-min epoch were more essential for the partitioning of the two groups than the features from whole hours. Our statistical evaluation showed that all the features evidenced significantly different mean values, with the exception of the low-level activity ratio, which was

© Schattauer 2010

H. J. Kam et al.: Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity

lower than 1.0 G. This is because moving down the branches of a decision-tree, the sizes of the subjects differed from the very beginning. We began with 142 subjects at the uppermost node, who were distributed over a variety of downstream branches while they passed through the conditional rudders at the nodes. The most distinctive activity feature in that situation might differ from that of the entire subject-base. In ▶Table 3, all the selected features were compared among the entire subject population; therefore, some activity features positioned at the lower parts of the trees could evidence insignificant results. However, the valid features utilized in tree construction were consistent with those of earlier studies, and were quantitative.

The result that there were distinctive features in activity that could statistically significantly differentiate ADHD and non-ADHD groups corroborates a common finding from other previously published works [18-22]: namely, activity between patients with/without ADHD differs significantly, as measured by actigraphs. The preexisting accelerometers for actigraph measurements have been previously utilized for long-term observations of objects' daily activities by offering counts or summaries of activity that exceed a pre-defined temporal threshold within a given period. On the contrary, the actigraphs utilized in this study provided high-resolution raw data on activities, thus allowing raw data to be processed into values within specific ranges of activity, and to make use of valuable features. Owing to the dimensional differences in the data, a direct comparison of our results with those of previous studies is difficult. However, two specified activity differences can be restated: the difference in high-level activity between two groups as reported by Teicher et al. [5] and the higher separation capability with higher feature dimensions [25].

From an in-depth view of the selected features, not only mere hyperactivity among children's manifold activities, but also variances in that activity, were selected as important features for group separation. The intensity and variability of activities may be critically important features for ADHD monitoring, relative to the mere counts of activities as in the previous studies; this was also corroborated by the results of another previous study [22]. These variances reflect instability of activity, and may also be reflective of other properties, including inattention, impulsivity, or hyperactivity: That is, the identification and distinction of patients with ADHD from others depends heavily on whether activity is maintained and to what extent, in addition to the amount of activity; additionally, those characteristics were reflected in the actigraphic data.

5.3 Measuring Environment for Children's Activity

The activity data were acquired on consecutive days with no alteration or disruption of the pre-established curriculum or school timetable (and each class); thus, the sequence and times of class courses among the children were not identical. In this study, we attempted to characterize the specific activity features among the children with ADHD that are conducted under natural living conditions rather than in hypothetical or systematically restricted ones, and to construct model for ADHD screening; therefore, measuring activities during identical tasks may be counter to the original objectives of our study. However, carefully designed measuring environments – such as standardized teaching methodologies and content, or emphasis on the classes with major subjects – can reduce more profoundly the dependence between the measurements and tasks.

5.4 Limitations

One major limitation of this study was that the group separations were based on the primary results of ADHD diagnosis. No sub-type classifications were applied in this study. Hence, it was not possible to completely dismiss the possibility of inequality or bias regarding subtypes [4, 34] and gender [1, 19]. In this study, we constructed the high-risk group using a preexisting questionnaire as the first step in obtaining a "gold standard" for modeling. Then, children in the high-risk group were interviewed and clinically diagnosed by experienced child psychiatrists. Because all of the other children-who comprised a low-risk group that was excluded from diagnosis confirmations-were put together into the non-ADHD group, it is possible that some of the patients with ADHD who were not filtered out by the questionnaire might have been misdiagnosed as non-ADHD individuals. Potential mis-grouping was, therefore, one of the limitations of this study.

© Schattauer 2010

H. J. Kam et al.: Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity

Additionally, this study employed a relatively small sample numbers. Because this research was conducted as a community-based study, the relatively low prevalence of ADHD (3-9%) permitted the inclusion of only a small number of children diagnosed with ADHD. Nevertheless, the prevalence of the object group in this study was 7.04% – well within the 3-9% range of established prevalence. Furthermore, when we set up the high-risk group, the cut-off points were 60 points and the upper 10% of K-CBCL and C-ARS, thus ensuring sufficiently high selectivity. Therefore, the consequences of regarding the un-diagnosed low-risk group as controls (non-ADHD) might be insignificant.

Another limitation of this study involved the lack of coverage information for the proposed screening model. Because the ADHD samples applied to the decision-trees were only from the high-risk group whose diagnoses were confirmed by questionnaires, there was no opportunity to compare the coverage between questionnaires and the screening model for ADHD diagnosis separately with unidentified subjects. We are currently attempting to compare simultaneously the coverage of two screening methodologies-questionnaires and the decision-tree-based screening model-to new unidentified subjects in a new experimental setting.

For the reasons enumerated above, caution should be exercised in treating the resultant screening features as clinically factual; still, the basic merit of this study was that it was conducted under natural conditions rather than under hypothetical or systematically restricted conditions. Additionally, clearly differentiated activity patterns were detected between children with/without ADHD which might have been ambiguous on questionnaires or when measured via the traditional activity features; thus, our methodology constitutes a new ADHD monitoring scheme, featuring new activity features, can be monitored over a broad range of activities.

6. Conclusion

The screening of patients with ADHD can be accomplished by making subject children wear a watch-like sensor for several hours while they attend school. The valid features used in tree construction are consistent with those reported in earlier studies, and the variance of the 1-min epoch mean and the mean and variance of high activity (1.8 G and 3.2 G) were the principal factors in group separation. Therefore, the ADHD screening model proposed herein is sufficiently objective and quantitative to be considered high-yield and applicable to a large number of children, without requiring vast quantities of time and manpower.

Clinical Relevance
Latent ADHD is the most common childhood neurobehavioral disorder that requires early detection and treatment, and a screening test applicable to many children at once is clearly worth a great deal from a public health perspective.

Conflict of Interest
All of the authors of this work hereby declare that there were no conflicts of interest or any other relationships that could have inappropriately influenced this study.

Human Research / IRB section
The subject children's parents provided signed statements of informed consent. This study was reviewed and is in compliance with the Declaration of Helsinki, and approved by the Institutional Review Board of the Ajou University Medical Center (AJIRB-CRO-07-130).
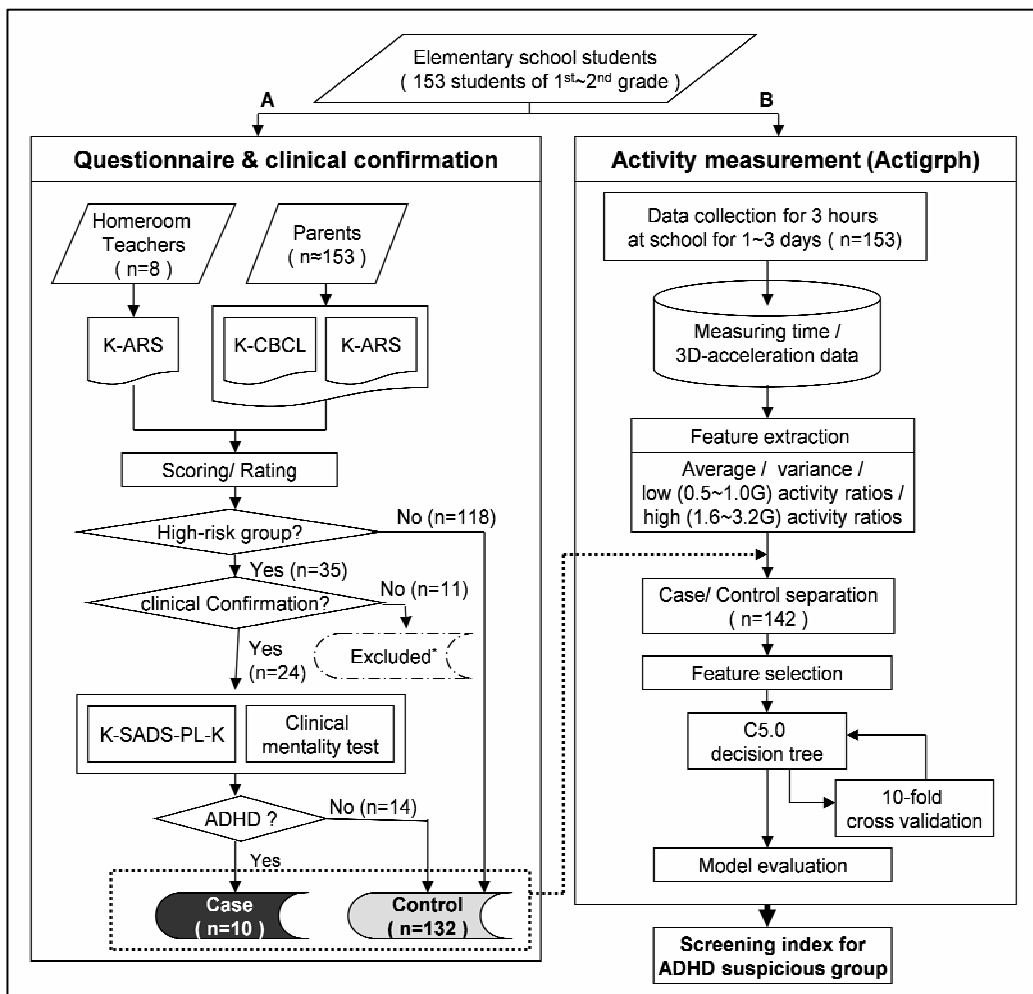
Fig. 1 Overview of the process-flow for ADHD screening model construction: (A) General process for screening & diagnosis of ADHD and (B) process to build screening models for ADHD with activity features (* Dropped from further clinical confirmation; K-ARS: Korean ADHD Rating Scale-IV; K-CBCL: Korean Child Behavior Checklist; K-SADS-PL-K: Kiddie Schedule for Affective Disorder and Schizophrenia-Present and Lifetime Version-Korean Version
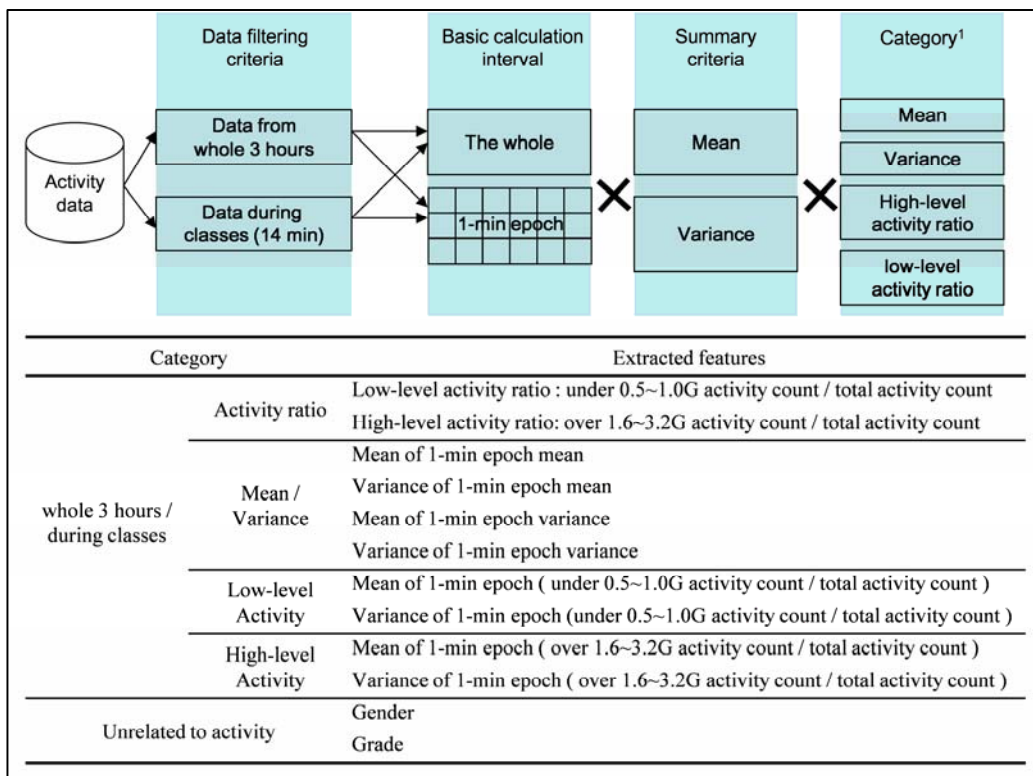
Category		Extracted features
whole 3 hours / during classes	Activity ratio	Low-level activity ratio : under 0.5~1.0G activity count / total activity count
		High-level activity ratio: over 1.6~3.2G activity count / total activity count
	Mean / Variance	Mean of 1-min epoch mean
		Variance of 1-min epoch mean
		Mean of 1-min epoch variance
		Variance of 1-min epoch variance
	Low-level Activity	Mean of 1-min epoch (under 0.5~1.0G activity count / total activity count)
		Variance of 1-min epoch (under 0.5~1.0G activity count / total activity count)
	High-level Activity	Mean of 1-min epoch (over 1.6~3.2G activity count / total activity count)
		Variance of 1-min epoch (over 1.6~3.2G activity count / total activity count)
Unrelated to activity		Gender
		Grade

Fig. 2 The designed feature extraction process from activity data (1 High-level activity increased at 0.2 G interval from 1.6 G, and low-level activity decreased at 0.1 G start from 1.0 G)
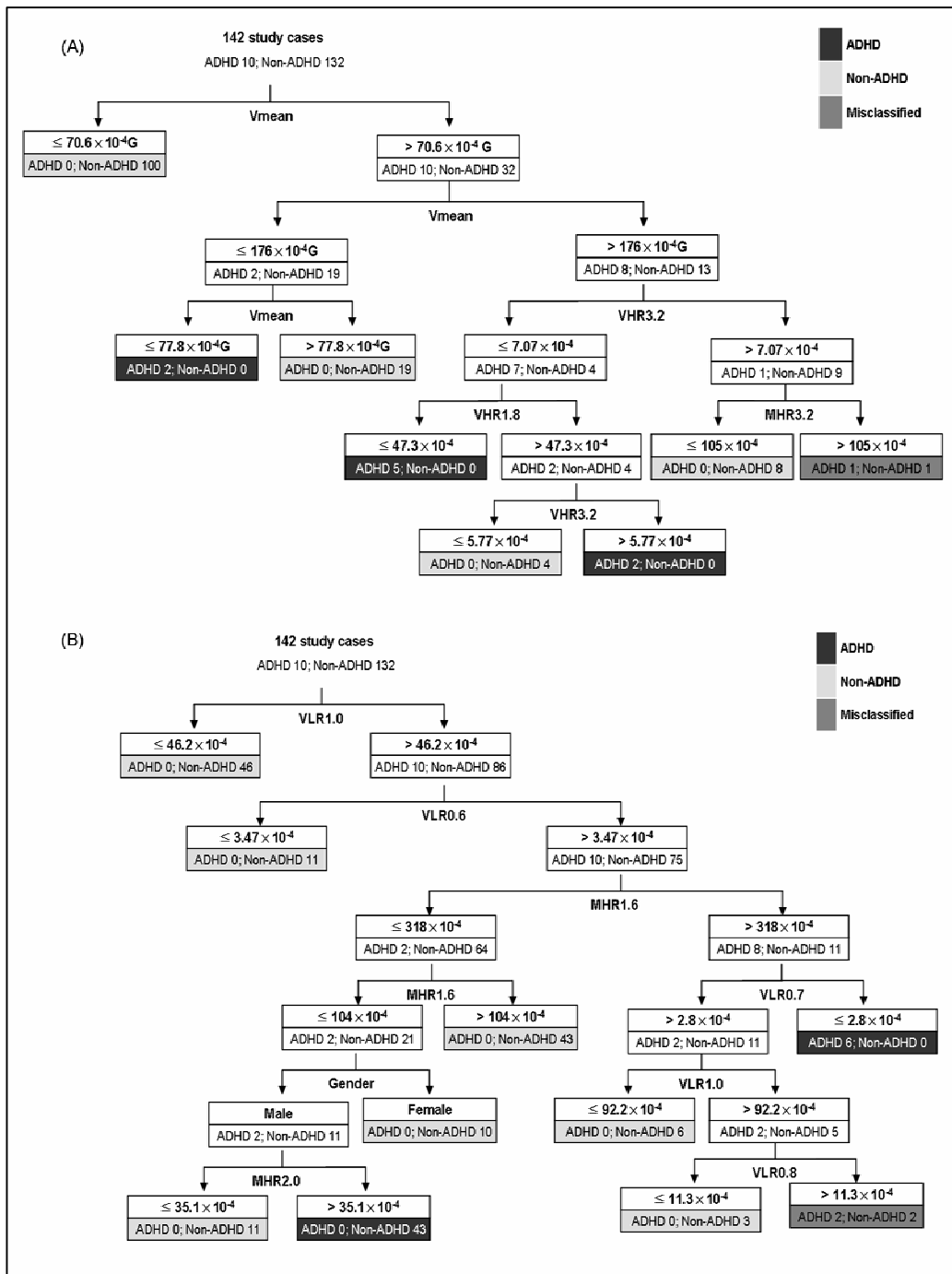
Fig. 3 Constructed decision trees for screening ADHD: (A) Model A (class+playtime) and (B) Model B (class hours) (G = gravity acceleration; Vmean = variance of mean; VHR1.8/VHR3.2 = variance of high-level activity ratio (>1.8G and >3.2G, respectively); VHR1.8 = variance of high-level activity ratio (>1.8G); MHR1.6/MHR2.0/MHR3.2 = mean of high-level activity ratio (>1.6G, >2.0G and >3.2G, respectively); VLR0.6/ VLR0.7/ VLR0.8/ VLR1.0 = variance of low-level activity ratio (<0.6G, <0.7G, <0.8G and <1.0G, respectively)

Table 1 Demographic and clinical characteristics of children with/without ADHD.

		ADHD (n = 10)	Non-ADHD (n = 132)	Total (n = 142)
Gender	Male	8 (5.6%)	63 (44.4%)	71(50.0)
	Female	2 (1.4%)	69 (48.6%)	71 (50.0)
Age, yr		7.20 ± 0.63	7.46 ± 0.62	7.44 ± 0.62
K-ARS. Attention deficit		12.90 ± 6.15	2.60 ± 3.60	3.37 ± 4.69
	Hyper activity	14.60 ± 7.66	2.48 ± 4.19	3.38 ±5.51
	Total scores	27.50 ± 13.77	5.07 ± 7.44	6.75 ± 9.95
K-CBCL, syndrome scales	Withdrawn	52.30 ± 11.30	46.95 ± 9.00	47.33 ± 9.23
	Somatic Complaints	50.39 ± 7.93	49.78 ± 10.54	49.82 ± 10.36
	Anxious/Depressed	56.80 ± 17.61	49.51 ± 9.37	50.03 ± 10.24
	Social Problems	59.77 ± 17.11	49.71 ± 8.71	50.41 ± 9.79
	Attention Problems	61.20 ± 13.34	48.82 ± 8.77	49.69 ± 9.64
	Delinquent Behavior	53.82 ± 9.77	49.34 ± 9.51	49.65 ± 9.56
	Aggressive behavior	62.54 ± 15.51	49.44 ± 9.10	50.36 ± 10.18
K-CBCL.	Internalizing problems	52.30 ± 11.30	46.95 ± 9.00	47.33 ± 9.23
	Externalizing problems	58.20 ± 9.96	48.15 ± 9.13	48.85 ± 9.51
	Total behavior problems	56.60 ± 12.00	46.42 ± 9.37	47.14 ± 9.88

All values, except gender row, are mean ± SD; K-ARS = Korean ADHD Rating Scale-IV; K-CBCL = Korean Child Behavior Checklist

Table 2 Demographic features of diagnosed 24 children and their parents.

		ADHD (n = 10)	Non-ADHD with OPP(s) (n = 5)	Normal (n = 9)
Gender	**Male**	8 (80%)	4 (80%)	3 (33%)
	Female	2 (20%)	1 (20%)	6 (67%)
Age (yr, mean ± SD)		7.20 ± 0.63	7.40 ± 0.55	7.44 ± 0.53
Family structure and living with both parents		10 (100%)	5 (100%)	9 (100%)
Former clinical consultation		1 (14%)	-	-
Clinical diagnosis[*]	**ADHD**	10 (100%)	-	-
	Emotional problems	-	5 (100%)	-
	Tic disorder	-	1 (20%)	-
Medication use (before activity measurement)		-	-	-
Parents – Monthly income ($)	**860 – 1,721**	1 (10%)	-	-
	1,722 – 2,581	6 (60%)	2 (40%)	6 (67%)
	2,582 – 3,442	1 (10%)	2 (40%)	1 (11%)
	3,443 – 4,303	1 (10%)	-	-
	≥ 4,304	1 (10%)	1 (20%)	2 (22%)
Parents – Education Father	**>12 years**	5 (50%)	-	6 (67%)
	9-12 years	5 (50%)	5 (100%)	3 (33%)
Parents – Education Mother	**>12 years**	3 (30%)	-	2 (22%)
	9-12 years	7 (70%)	5 (100%)	7 (78%)

[*] There was a condition with two co-morbidities for one child; OPP = Other psychiatric problem.

Table 3 Comparison of activity levels with regard to selected features (with 1-min epoch).

Type	Decision Node Feature	ADHD [Unit: G×10⁻⁵] (mean ± SD)	Non-ADHD [Unit: G×10⁻⁵] (mean ± SD)	*P* value*
Model A (Class+playtime)	Vmean	1939.74 ± 823.43	580.70 ± 800.27	<0.001
	VHR1.8	410.91 ± 228.21	138.01 ± 203.95	<0.001
	VHR3.2	51.56 ± 35.68	13.48 ± 26.99	<0.001
	MHR3.2	719.90 ± 354.94	261.45 ± 294.69	<0.001
Model B (Class hours)	VLR0.6	12.42 ± 9.68	10.15 ± 15.33	0.023
	VLR0.7	32.79 ± 21.16	24.61 ± 33.52	0.020
	VLR0.8	98.39 ± 80.98	59.85 ± 69.66	0.012
	VLR1.0	1900.51 ± 2026.78	1228.66 ± 1365.13	0.084
	MHR1.6	3745.62 ± 1596.46	2046.78 ± 1627.28	0.006
	MHR2.0	1882.28 ± 915.79	906.76 ± 895.57	0.003

*Mann-Whitney U-test; G = Gravity acceleration; Vmean = Variance of mean; VHR 1.8/VHR 3.2 = Variance of high-level activity ratio (>1.8 G and >3.2 G, respectively); VLR 0.6/ VLR 0.7/ VLR 0.8/ VLR 1.0 = Variance of low-level activity ratio (<0.6 G, <0.7 G, <0.8 G and <1.0 G, respectively); MHR 1.6/ MHR 2.0/ MHR 3.2 = Mean of high-level activity ratio (>1.6 G, >2.0 G and >3.2 G, respectively).

Table 4 Model evaluation for constructed decision trees.

	Accuracy (%)	Sensitivity	Specificity	PPV		NPV		Likelihood ratio†	AUC
				(3%)*	(7%)*	(3%)*	(7%)*		
Model A (class+play-time)	99.30	1.000	0.992	0.803	0.909	1.000	1.000	(+) 132 (-) 0	0.9996
Model B (class hours)	98.59	1.000	0.985	0.671	0.832	1.000	1.000	+) 66	0.9985

*Assumed prevalence; † (+) tells how much the odds of the disease increase when a test is positive; (-) tells how much the odds of the disease decrease when a test is negative

PPV = Positive predictive value; NPV = Negative predictive value; AUC = Area under ROC curve.

References

1. Teicher MH, Ito Y, Glod CA, Barber NI. Objective measurement of hyperactivity and attentional problems in ADHD. J Am Acad Child Adolesc Psychiatry 1996; 35(3): 334-342.
2. Kofler MJ, Rapport MD, Alderson RM. Quantifying ADHD classroom inattentiveness, its moderators, and variability: a meta-analytic review. J Child Psychol Psychiatry 2008; 49(1): 59-69.
3. Goodyear P, Hynd GW. Attention-deficit disorder with (ADD/H) and without (ADD/WO) hyperactivity: behavioral and neuropsychological differentiation. J Clin Psychol 1992; 21(3): 273-305.
4. Dane AV, Schachar RJ, Tannock R. Does actigraphy differentiate ADHD subtypes in a clinical research setting? J Am Acad Child Adolesc Psychiatry 2000; 39(6): 752-760.
5. Teicher MH. Actigraphy and motion analysis: new tools for psychiatry. Harv Rev Psychiatry 1995; 3(1): 18-35.
6. Swanson JM, Gupta S, Williams L, Agler D, Lerner M, Wigal S. Efficacy of a new pattern of delivery of methylphenidate for the treatment of ADHD: Effects on activity level in the classroom and on the playground. J Am Acad Child Adolesc Psychiatry 2002; 41(11): 1306-1314.
7. Pärkkä J, Ermes M, Korpipää P, Mäntyjärvi J, Peltola J, Korhonen I. Activity classification using realistic data from wearable sensors. IEEE Trans Inf Technol Biomed 2006; 10(1): 119-128.
8. Choudhury T, Philipose M, Wyatt D, Lester J. Towards activity databases: using sensors and statistical models to summarize people's lives. IEEE Data Eng Bull 2006; 29(1): 49-58.
9. Hong JH, Kim NJ, Cha EJ, Lee TS. Classification technique of human motion context based on wireless sensor network. Conf Proc IEEE Eng Med Biol Soc 2005; 5: 5201-5202.
10. Liao L, Fox D, Kautz H. Location-based activity recognition using relational markov networks. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) 2005; 30 July - 5 August 2005; Edinburgh, Scotland; 2005: 773.
11. Burchfield TR, Venkatesan S. Accelerometer-based human abnormal movement detection in wireless sensor networks. In: 1st International Workshop on Systems and Networking support for Healthcare and Assisted living Environments (HealthNet's 07); 11 June 2007; San Juan, Puerto Rico, USA; 2007: 67-69.
12. Zhang T, Wang J, Liu P, Hou J. Fall detection by embedding an accelerometer in cellphone and using KFD algorithm. Intl. J Comp Sci Net Sec 2006; 6(10): 277-284.
13. Chen J, Kwong K, Chang D, Luk J, Bajcsy R. Wearable sensors for reliable fall detection. Conf Proc IEEE Eng Med Biol Soc 2005; 4: 3551-3554.
14. Laerhoven KV, Gellersen HW, Malliaris YG. Long-term activity monitoring with a wearable sensor node. In: International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06); 3-5 April 2006; Cambridge, Massachusetts, USA; 2006: 170-174.
15. Tuisku K, Virkkunen M, Holi M, Lauerma H, Naukkarinen H, Rimon R, et al. Antisocial violent offenders with attention deficit hyperactivity disorder demonstrate akathisia-like hyperactivity in three-channel actometry. J Neuropsychiatry Clin Neurosci 2003; 15(2): 194-199.
16. Tuisku K, Lauerma H, Holi M, Markkula J, Rimon R. Measuring neuroleptic-induced akathisia by three-channel actometry. Schizophr Res 1999; 40(2): 105-110.
17. Van Someren EJW, Pticek MD, Speelman JD, Schuurman PR, Esselink R, Swaab DF. New actigraph for long-term tremor recording. Mov disord 2006; 21(8): 1136-1143.
18. Halperin JM, Matier K, Bedi G, Sharma V, Newcorn JH. Specificity of inattention, impulsivity, and hyperactivity to the diagnosis of attention-deficit hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 1992; 31(2): 190-196.
19. Porrino LJ, Rapoport JL, Behar D, Sceery W, Ismond DR, Bunney WE Jr. A naturalistic assessment of the motor activity of hyperactive boys. I. Comparison with normal controls. Arch Gen Psychiatry 1983; 40(6): 681-687.
20. Halperin JM, Newcorn JH, Matier K, Sharma V, KMS, Mckay KE, Schwartz S. Discriminant validity of attention-deficit hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 1993; 32(5): 1038-1043.
21. Tsujii N, Okada A, Kaku R, Kuriki N, Hanada K, Matsuo J, et al. Association between activity level and situational factor in children with attention deficit/hyperactivity disorder in elementary school. Psychiatry Clin Neurosci 2007; 61(2): 181-185.
22. Wood AC, Asherson P, Rijsdijk F, Kuntsi J. Is overactivity a core feature in ADHD? Familial and receiver operating characteristic curve analysis of mechanically assessed activity level. J Am Acad Child Adolesc Psychiatry 2009; 48(10): 1023-1030.
23. Dupaul GJ, Power TJ, Anastopoulos AD, Reid R. ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation. New York, NY: The Guilford Press; 1998.
24. Kim YS, Cheon, KA, Kim BN, Chang SA, Yoo HJ, Kim JW, et al. The reliability and validity of Kiddie-Schedule for affective disorders and schizophrenia-present and lifetime version-Korean version (K-SADS-PL-K). Yonsei Med J 2004; 45(1): 81-89.

25. Teicher MH, McGreenery CE, Ohashi K. Actigraph assessment of rest-activity disturbances in psychiatric disorders. In: Psychosomatic Medicine - Proceedings of the 18th World Congress on Psychosomatic Medicine; 21-26 August 2005; Kobe, Japan; 2006; 32-37.

26. Yun C, Shin D, Jo H, Yang J, Kim S. An experimental study on feature subset selection methods. In: 7th IEEE International Conference on Computer and Information Technology (CIT); 16-19 October 2007; Fukushima, Japan; 2007: 77-82.

27. Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. Academic Press (CA): Morgan Kaufmarm Publishers, Inc. 2001; 284-291.

28. Tan P-N, Steinbach M, Kumur V. Introduction to data mining. 1st ed. Addison-Wesley (MA): Pearson Education, Inc. 2006.

29. Matthew NA, Sajjan S. Comparative analysis of serial decision-tree classification algorithms. International Journal of Computer Science and Security 2009; 3(3): 230-240.

30. Rulequest research 2009 [Internet]. Australia: RuleQuest Research Pty Ltd; c1997-2008 [updated 2009 Nov; cited 2010 Jul 23]. Is see5/C5.0 setter than C4.5? Available from: http://www.rulequest.com/see5-comparison.html

31. Quinlan JR. Programs for Machine Learning. Morgan Kaufmann Publishers, Inc. 1993.

32. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. Intelligent Data Analysis 2002; 6(5): 429-449.

33. Kim JW, Park KH, Cheon KA, Kim BN, Cho SC, Hong KEM. The child behavior checklist together with the ADHD rating scale can diagnose ADHD in Korean community-based samples. Can J Psychiatry 2005; 50: 802-805.

34. Dabkowska MM, Pracka D, Pracki T. Does actigraphy differentiate ADHD subtypes? Eur Psychiatry 2007; 22 (Suppl. 1): S319.