Objective ADHD diagnosis using Convolutional Neural Networks over Daily-Life Activity Records

Patricia Amado-Caballero, Pablo Casaseca-de-la-Higuera, Susana Alberola-López, Jesús María Andrés-de-Llano, José Antonio López Villalobos, José Ramón Garmendia-Leiza, Carlos Alberola-López

***Abstract*— Attention Deficit/Hyperactivity Disorder (ADHD) is the most common neurobehavioral disorder in children and adolescents. However, its etiology is still unknown, and this hinders the existence of reliable, fast and inexpensive standard diagnostic methods. *Objective:* This paper proposes an end-to-end methodology for automatic diagnosis of the combined type of ADHD. *Methods:* Diagnosis is based on the analysis of 24 hour-long activity records using Convolutional Neural Networks to classify spectrograms of activity windows. *Results:* We achieve up to 97.62% average sensitivity, 99.52% specificity and AUC values over 99%. Overall, our figures overcome those obtained by actigraphy-based methods reported in the literature as well as others based on more expensive (and not so convenient) acquisition methods. *Conclusion:* These results reinforce the idea that combining deep learning techniques together with actimetry can lead to a robust and efficient system for objective ADHD diagnosis. *Significance:* Reliance on simple activity measurements leads to an inexpensive and non-invasive objective diagnostic method, which can be easily implemented with daily devices.**

***Index Terms*— ADHD, actigraphy, Deep Learning, Convolutional Neural Network (CNN)**

I. Introduction

ATTENTION-deficit/hyperactivity disorder (ADHD) is one of the most common neurobehavioral disorder in school age population [1]. In this cohort, its prevalence is about 7.2% [2] depending on the diagnostic criteria and the studied population (for instance, in Spain prevalence is 6.8% [3]).

Manuscript received 11/07/19; revised 12/11/19; accepted 29/12/19. Date of publication ZZZZ; date of current version 01/02/20. This work was supported in part by the European Regional Development Fund (ERDF-FEDER) under Grant TEC2017-82408-R, in part by the Instituto de Competitividad Empresarial (Junta de Castilla y León, Spain under Grants 04/18/VA/0008 and 04/18/VA/0013), and in part by the Scottish Funding Council through CENSIS Innovation Centre under grant CAF-0040. (Corresponding author: P. Casaseca-de-la-Higuera. E-mail: casaseca@lpi.tel.uva.es).

P. Amado-Caballero, C. Alberola-López and P. Casaseca-de-la-Higuera are with UIC Laboratorio de Procesado de Imagen. ETSI Telecomunicación, Dpto. Teoría de la Señal y Comunicaciones e Ingeniería Telemática, 47011 Valladolid, Spain.

P. Casaseca-de-la-Higuera is also with the School of Computing, Engineering, and Physical Sciences, University of the West of Scotland, Paisley Campus, High Street, Paisley, PA1 2BE, Scotland, United Kingdom

S. Alberola-López is with Centro de Salud Jardinillos. 34001 Palencia, Spain

J. M. Andrés-de-Llano, J. A. López Villalobos, J. R. Garmendia-Leiza are with Complejo Asistencial Universitario de Palencia. 34005 Palencia, Spain

This diagnostic protocol is currently defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5 [4]) . However, even though the treatment of ADHD is well-defined for the known types —inattentive, hyperactive-impulsivity and combined— it's not easy either to find a specific etiology or to define an objective diagnostic method.

ADHD is a disorder defined by a persistent pattern of inattention and/or hyperactivity-impulsivity, which is more frequent and severe than that usually observed in subjects of a similar level of development [4]. Deciding if this pattern actually constitutes an anomaly with respect to the normal development of a child is a task that falls entirely on the observations of parents and teachers as well as on the interpretation that specialists make out of these observations; actually results may differ whether parents, teachers or both are taken as source of information [5].

In light of the above, there is need of reliable and objective diagnostic procedures for clinical evaluation of ADHD. Over the years, some studies have tried to find associations between physical signals and the disorder. MRI (Magnetic resonance imaging) is frequently used for mental disorders; in the context of our problem, different studies used this technique to obtain images of the brain to detect differences between patients diagnosed with ADHD and control patients. The use of volumetric images of the brain [6] or the study of iron levels [7] are two examples of attemtps to make diagnosis objective. In some studies on ADHD [8] functional magnetic resonance (fMRI) has also been used to study active areas of the brain in older people. Differences between healthy and ADHD patients have also been observed by means of EEG (electroencephalography) [9]–[12]. However, despite the efficacy of these methods, both their high costs as well as the very nature of the pathology question their actual reliability as effective methods [13].

Some investigations support the hypothesis of the relationship between ADHD and sleep cycles and present this test as a possible objective diagnostic method for this pathology [14] . These studies have used Polysomnography (PSG) EEG, electrooculography (EOG), electromiography (EMG), electrocardiography (ECG), respiratory signals and pulsioximetry (POX) [15], [16]. Actigraphy measurements have also been reported as a useful method for the diagnostic of ADHD [17].

Nowadays, hardware advances, mainly with the use of GPUs and parallel calculations, have triggered a paradigm shift in scientific research, making it possible to create artificial intelligence systems that process and learn from vast amounts of data (i.e., the deep learning paradigm). In recent works, deep learning algorithms have been used for feature extraction in sleep patterns studies [18]–[20]. Some authors have used deep learning with the data obtained from accelerometers in order to find patterns in ADHD children [21].

The purpose of this paper is to design, train and deploy a highly accurate expert system based on a Convolutional Neural Network (CNN) that is able to diagnose combined ADHD out of the a 24-hour-long actigraphic record of a child in a regular school day. This is an affordable and non-invasive measure that does not interfere with the child's daily life. As our forthcoming analysis of the state of the art reveals (see section II), most of the attempts to create expert systems are carried out in controlled experiments (including [21]) or, at least, describe targeted experiences with somewhat limited patient cohorts.

The paper is structured as follows; section II describes the main contributions in the field; they are schematically summarized in Tables I–III for quick reference. Section III describes the materials employed and the methodology applied. Results are shown in section IV, and are then discussed and compared with previous proposals in section V. Concluding remarks are summarized in section VI.

II. State of Art

Depending on the source of information, ADHD diagnosis could be subjective or objective. It typically has been diagnosed from parental and teacher reports as well as medical questionaries, i.e., subjectively. Over the years, several studies have proposed different diagnosis methods.

As for the the methodologies of interest in this paper, attempts for objective diagnostic procedures can be classified according to:

- Non-actimetric methods: They propose the use of MRI and EEG for diagnosis. In the first case, several studies have achieved positive results in the classification of subtypes of ADHD. In the case of EEG, promising results have been obtained regarding the differentiation of patients. As previously stated, they are expensive methods and although interesting results have been reported, the nature of the pathology questions whether they are a cost-effective diagnostic option.
- Actimetric methods: Their main difference with respect to other methods is their lower cost. Described results show effectiveness since fairly good results are obtained.

Tables I–III present a summary of the most relevant —to the best of our knowledge— ADHD studies in the recent past (since 2011). Earlier contributions can be found in [17]. Table I focuses on non-actimetric studies whereas Tables II and III respectively focus on rhythmometry and measurements of activity. The tables have been structured in terms of Materials, Methods and Results & Conclusions in an attempt to make them self-contained. Results are those reported by the authors in the cited papers. In what follows, we focus on those studies that use different types of machine learning/deep learning algorithms.

The study presented by Riaz et al. [22] used a Support Vector Machine (SVM) to classify data obtained from the repository *NeuroBureau ADHD-200* [35] . These data consist of a specific type of fMRI known as *Resting State fMRI* (rsfMRI or R-fMRI) that evaluates regional interactions occurring when an explicit task is not being performed. The achieved classification accuracy was 0.818. Oztoprak et al. [12] analyzed ERP (Event-related potentials) from EEG signals during a Stroop-type task to identify ADHD patients. They employed a SVM classifier with recursive feature elimination fed with high resolution time-frequency domain features. They achieved 100% accuracy using a test group of 10 subjects, whereas the overall accuracy over the train set was 0.995. Sun et al. [23] used radiomics for ADHD diagnosis using MRI. This method relies on the extraction of a large amount of features from medical imaging, to obtain useful information for diagnosis. The achieved accuracy was 0.737 using cross-validation with random forests.

Relevant publications using machine learning over actimetry can also be mentioned. The study by Muñoz et al. [21] used two accelerometers placed on the wrist and ankle respectively. Activity data along 6 school hours in a group of small children (22 patients, 11 ADHD and 11 healthy) was analysed using a CNN. The obtained accuracy from the wrist device was 0.8750, whereas the one from the ankle was 0.9375. Sensitivity values were 0.6 and 0.8, respectively. In [32], Mahony et al. used gyroscopes for motion characterization as well as accelerometers. Classification of the best performing features using SVMs, yielded 0.9512 accuracy, 0.9444 sensitivity, and 0.9565 specificity.

According to this analysis, our proposal can be grounded on these three main ideas:

- Actimetry seems more convenient for the study of ADHD, since it guarantees an objective and non-intrusive diagnostic method, of lower cost and more comfortable for the patient.
- The methods based on actimetry in recent years are presenting results comparable to those presented in previous studies using MRI and EEG.
- Previous results obtained with actigraphy show good figures but cohorts seem somewhat limited and/or experiments require a controlled scenario.

III. Materials and methods

A. Materials

Our subject group includes children with ages between 6 and 15 years who were monitored in regular daily activity for a period of approximately 24 hours. The group was composed of 148 subjects, of which 73 had been diagnosed as ADHD combined type according to the DSM-5 criteria. None of them had taken medication at the time of the study. The other 75 subjects are healthy subjects who make up the control group. Hereinafter, we will denote by *cases* the group of ADHD patients and by *controls* the group of healthy subjects. The study was carried out in accordance with the Declaration of Helsinki and was approved by the *Área de Salud de Palencia* Research Ethics Committee (REC number: 15/SS/0234). Subjects provided their informed consent before the recordings.

TABLE I: State of the art on ADHD assessment based on non-actimetric sources. Results are those reported by the authors.

Medical Imaging

Ref.	Materials:	Methods:	Results and Conclusion:
[22]	*NeuroBureau ADHD-200*	*Resting State fMRI* (rsfMRI o R-fMRI)	Accuracy = 0.818, Sensitivity=1, Specificity = 0.75
[?], [23]	83 paired children by sex and age diagnosed as non-treatment (40 with innatentive ADHD and 43 with combined ADHD) and 87 control subjects	Anatomical MRI and diffusion tensor Images	No overall differences were found between the children with ADHD and the control subjects in the total brain volume or in the total volume of gray and white matter. Accuracy = 0.737 (discrimination of patients) and Accuracy = 0.801 (different subtypes)

Biomedical Signals

Ref.	Materials:	Methods:	Results and Conclusion:
[12]	Between 37 and 44 children in ADHD group and 32 and 38 children on healthy group.	ERP measurements in the EEG signals during the performance of a Stroop type task.	Accuracy = 0.995 (training set) Accuracy = 1 (10 test subjects)

TABLE II: State of the art on ADHD assessment through circadian activity rhythm. Results are those reported by the authors.

Rythmometry

Ref.	Materials:	Methods:	Results and Conclusion:
[24]	9 normally developed children aged 6–11 years old registered during 6 days with actigraphy.	Hierarchical multivariate regression between actigraphy registry and Strength and Difficulties Questionaire (SDQ) to evaluate the effect of sleep on SDQ.	Sleep accounts for 18% of the variance in conduct problems. Only the actual time of sleep is significant within the model ($p < 0.05$). Children that sleep 1 hour less than the average, are in risk of behavioral troubles.
[25]	37 children affected of combined kind of ADHD were registered with actigraphy during 3 days.	Comparison between actigraphy measures and results of the ADHD–RS (ADHD Rating Scale), filled by parents during 6 months.	The peak time of the COSINOR method correlates with inattention symptoms; specifically $\rho = 0.349$, with significance $p = 0.057$.
[26]	10 adults diagnosed of ADHD (combined and intent subtypes) treated with Methylphenidate (MPD). 1st week without MFD, and 2nd, 3rd and 4th week with increasing dose of MPD.	Actigraphy registries were recorded for 4 weeks both at daytime and at night. Saliva melatonin was also measured every day. Patients also filled a ADHD–RS questionary. Authors do not specify what actigraphic variables are measured.	There is a correlation between MFD dose and melatonin level in saliva and results in ADHD–RS, however these results were not found for the actigraphic analysis. Author states that the use of different actigraphy variables justify the discrepancy between this word and literature.

Actigraphic signals were acquired with the *ActiGraph GT3x* device [36], placed on the dominant wrist of each patient. This device measures the acceleration on each of the three Cartesian axes, registering a sample every second (operating frequency is $f_s = 1$ Hz). The aggregate signal $r = ||(x, y, z)|| = \sqrt{(x^2 + y^2 + z^2)}$ will be the measurement used as the input signal, i.e., a 1D signal is used per patient. The actual measurement is a *count* [37], as provided by the commercial device.

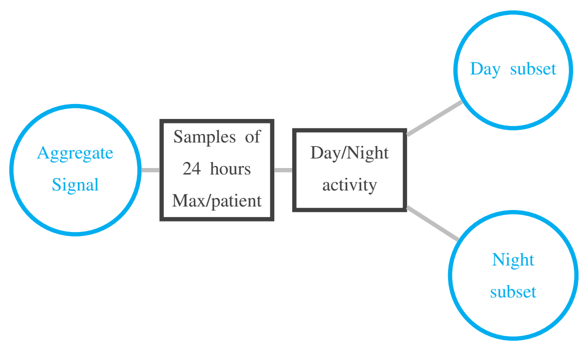

Fig. 1: Stages of the adequacy of the records of information, prior to preprocessing

B. Methods

We process the signals following the procedure shown in Fig. 1. Specifically, we adjust the number of samples up to a maximum of 24 hours per patient; taking into account that the actimeter worked at $f_s = 1$ Hz, this gives rise to a maximum of 86400 samples. Then, we divide the signal into two different subsets based on the activity periods (day time and night time activities) and they will feed two independent networks; we aim to identify differences between cases and controls on the basis of differences in both activity periods [38].

For each activity subset, we split the signal in windows with the same size. The window length is set to detect differences in more impulsive actions (i.e., with a short duration) and in more steady patterns (i.e., with a longer duration), following [17]; an in-between behaviour will also be inspected. Specifically, for each patient we will analyze windows of short-term activity (60 seconds), medium-term activity (5 minutes) and long-term activity (30 minutes). We have also used 66.67% window overlap in each of the cases.

1) Preprocessing: CNNs are especially efficient in the detection of patterns in images so 1D signals have to be appropriately conditioned to feed those networks. Specifically, the use of spectrograms as input data to a CNN has been shown to be effective for signals obtained through an actigraphic signal sensor [39]. We adhere to this approach, with the following specificities:

- *Data normalization*: Each data set has been normalized by dividing by the maximum value encountered in the registry.
- *Identification of low activity windows*: Visual inspection of the signal record shows the presence of [in]activity periods in which the actimeter registers null or too small values; these low activity windows could bias the network or give rise to inefficient training. So, discarding them is advisable. To this end, we studied the power distribution within the training data and set a threshold of minimum power per window to avoid the abnormally high bins typically located on the left end of the distributions. Windows

TABLE III: State of the art on ADHD assessment through activity measurement. Results are those reported by the authors.

Free Activity 1/2: ***24 hours***

Ref.	Materials:	Methods:	Results and Conclusion:
[17]	31 children affected of ADHD (combined subtype) and 32 healthy children aged 6 years old registered during 24 hours with actigraphy.	Activity is analyzed through nonlinear methods: central tendency measure and symbolic dynamics. Features resulting from this analysis are combined to construct a classifier.	Best classifier reaches an accuracy level of 0.9048 (sensitivity = 0.9677, specificity = 0.8438, AUC≈0.95).
[27]	31 children affected of ADHD (combined subtype) and 32 healthy children aged 6 years old registered during 24 hours with actigraphy.	Activity is analyzed through a the central tendency measurement for vector signals. Features resulting from this analysis are combined to construct a classifier.	Best classifier reaches an accuracy level of 0.9206 (sensitivity = 0.9355, specificity = 0.9062, AUC=0.9516).

Free Activity 2/2: ***Sleep Quality Assessment***

Ref.	Materials:	Methods:	Results and Conclusion:
[28]	26 patients affected of ADHD and 56 normally developed children aged 7–11 years old were registered during 5 days with actigraphy.	Actigraphy sleep report and Multiple Sleep Latency Tests (MSLT) were obtained to assess both the daytime and nighttime sleep pattern.	Longer sleep latency and longer sleep restlessness (both measured through actigraphy) were positively correlated to longer mean sleep latency at the MSLT.
[29]	24 patients affected of ADHD and insomnia aged 19–to–65 years old were monitored through actigraphy for two weeks.	Both sleep quality and circadian rhythm were registered through actigraphy. A backward stepwise regression was carried out to identify the relationship between the parameters of both analysis.	ADHD symptoms correlate with delayed sleep time and increased sleepiness ($p < 0.05$).
[30]	41 children affected of ADHD (21 of inattentive subtype, 2 impulsive, and 18 combined), aged 6–to–13 years old, and 41 aged–paired controls (± 6 months), with and without psychiatric comorbidities.	Sleep quality assessment through actigraphy.	Only ADHD patients affected of psychiatric comorbidities showed statistically significant differences with respect to the control group. Longest sleep latency ($p < 0.001$), smaller total sleep time ($p < 0.001$) and sleep efficiency ($p < 0.01$), higher nocturnal motor activity ($p < 0.001$) and wake after sleep onset ($p \simeq 0.05$).

Activity Performing a Specific Task

Ref.	Materials:	Methods:	Results and Conclusion:
[31]	20 ADHD patients and 15 healthy controls aged 18–24 years old.	Motor activity is evaluated when patients are performing different tasks demanding working memory.	ADHD patients have more movement than controls when they are developing tasks demanding working memory ($p = 0.034$).
[21]	22 patients, 11 ADHD and 11 healthy subjects	Two accelerometers on the wrist and ankle respectively to analyse data obtained in 6 school hours and a CNN to classify 5s-windows of 2D (horizontal and vertical) activity patterns as ADHD or controls. Final diagnosis is obtained by a combined likelihood measure obtained from the CNN output.	Accuracy = 0.8570 , Sensitivity = 0.6 and Specificity = 1 for the wrist and Accuracy= 0.937, Sensitivity =0.8 y Specificity=1 for the ankle. No specifics on the differences between ADHD and controls activity patterns were provided.
[32]	19 ADHD patients and 24 healthy controls aged 6–11 years old.	Activity of subjects were monitored through two inertial movement sensors placed on waist and the non–dominant ankle during the visit to the psychiatrist (1 hour approximately). A set of features reported in literature were extracted in several scenarios	Best results are achieved for a SVM–based classifier with 10 features; accuracy = 95.12%, sensitivity = 94.44% and specificity = 96.65%.
[33]	11 ADHD (combined kind) patients and 11 healthy controls between the ages of 8 and 12 years old. Actigraphy device placed on their non dominant wrist and ankles.	Activity was measured while patients were performing several tasks demanding attention and control.	ADHD patients move significantly more than healthy controls for both kinds of tasks. No differences in activity level were observed between the inhibition and noninhibition experimental tasks for either group.
[34]	5 ADHD patients and 11 healthy controls aged 3–6 years old. Patients were recorded in their sleep through a video camera; in addition, PSG registries were simultaneously acquired to determine the sleep–stage at each moment.	Automatic analysis of video was performed to assess the gross body moments to further extract the gross body movement rate and the rest period.	ADHD patients move more during every stage of sleep, specially during the REM stage.

below this threshold were set aside. The threshold has been empirically set to 0.022, and it is common for both day and night time as well as for short, medium and long term activity windows.

- *Generation of spectrograms*: We pursue to create spectrogram images that are visually smooth, have no discontinuities and allow for a correct visualization of the areas with frequency peaks [39]. In our particular case, we will have a unique CNN architecture for the three activity terms (as described in Section III-B.2), which means that all the input images will have the same size regardless of the size of the activity term (short/medium/long) and period (day/night). Specifically, the spectrogram will consist of a 129×55 matrix. Table IV shows the parameters used; we create a matrix of V rows and k columns; columns are created out of the signal points within a sliding window of length V that overlaps with the previous window in O points; hence, window shift is $V - O$. For this parameter choice $k = \frac{L-V}{V-O} + 1 = 55$. Then the `spectrogram` function in Matlab [40] performs a DFT with length 129 on each column.

Signal length(L)	Sliding window Size (V)	Overlap size (O)
60	6	5
300	30	25
1800	180	150

TABLE IV: Window parameters selection in seconds (and, equivalently, samples) for the spectrogram image generation.

2) CNN Architecture: The network used in the paper is depicted in Fig. 2; it consists of 3 convolutional layers with 32, 64 and 128 filters respectively, all followed by an activation layer (ReLU) and a normalization layer (BatchNormalization); the network also contains two pooling layers, a fully-Connected layer and a softmax layer. This architecture was achieved after an empirical validation stage; we started with a single convolutional layer and then the network was trained and validated. The process was repeated with an increasing number of convolutional layers until no significant improvements were obtained. A similar methodology was followed to select both the number and the dimensionality of the filters, where an appropriate balance between training duration and performance was sought. With this methodology, we came up with a solution that has also been reported by other authors as their design choice [41]–[44].

Random cross validation with ten different folds was used for training. Training was performed in two different ways (see Fig. 3):

- Training for each activity term: We have trained separately with images corresponding to the periods of short, medium and long-term activity, as previously defined.
- Training according to activity period: we have independently trained the network to find patterns in the signals pertaining to night or daytime activity.

Consequently, we have a unique network architecture but we end up with 6 different parameter settings (see Fig. 3). In each case, 70% of the patients constitute the training set and the rest the Test Set. Experiments have been run in Matlab, with the default parameters for the weights and learning rate, and 75 epochs in each training.

3) Final Classifier: The network in Fig. 2 is a binary classifier (Case/Control) for each input image (activity window). However, we pursue the overall classification of each patient, the record of which consists of large number of such images (windows). The final hard decision is accomplished as follows (see schemeatic of this pipeline in Fig. 4.:

- Select an activity term (short/medium/long). For this term, feed the network trained for day time activity with all the images that result from that period, following the procedure described in section III-B.2. Proceed similarly with the night activity subset.
- Calculate the fraction of spectrograms classified as Control and Case in the daytime interval. Repeat this calculation for the night time interval.
- For each patient, these two figures are used as the input to a two-dimensional binary classifier. This is where the final hard decision is made. To this end, we use a support vector machine (SVM) classifier with radial basis function (RBF) [45].
- As for performance evaluation, we repeat the same process in the 10 folds used in the training and for the different sets of window sizes.

IV. Results

In this section we first show some illustrative results on spectrograms. Then, we show classification performance figures as well as scatter diagrams of the samples used and its associated labels. This second part is twofold; specifically, we first report results with a classifier built on a $70-30\%$ train/test proportion. Then, and with the aim of checking classifier robustness, we repeat the experiments with a $20-80\%$ train/test proportion.

A. Spectrograms

Following the parameter selection shown in table IV we give rise to the different images used for the CNN input. Figs. 5, 6 and 7 respectively show examples of short-term, medium-term, and long-term spectrograms for one ADHD patient and one control. A high variability in the temporal patterns can be observed in all figures, since this depends on the specific activity the subject is carrying out on a particular moment. Regarding frequency patterns there is also high variability in the spectral content at different time stamps, both for the patient and control. We intend to illustrate with this particular example the difficulty of identifying a group-specific spectral signature by visual inspection. This highlights the need of a powerful feature extractor, such as our CNN, to automatically identify appropriate signatures so that differences between ADHD and controls can be obtained.

B. Global Patient Classification

Table V shows the performance of the classification system in terms of the following figures of merit, which have been calculated from the True and False positive (TP and FP) and negative (TN and FN) rates: 1) Upper part (from left to right)[1] $\Rightarrow$ accuracy, sensitivity, specificity, and area under the Receiver Operating Characteristic –ROC– curve (AUC); 2) Lower part (from left to right)[2] $\Rightarrow$ positive and negative predictive values (PPV and NPV) and likelihood ratios (LR+ and LR-). An ADHD patient correctly diagnosed as such is considered a TP whereas a control diagnosed as ADHD is a FP. A FN in classification is obtained for a non-diagnosed ADHD patient whereas a correctly identified control yields a TN.

To compare the performance of our spectrogram-based 2D CNN (CNN 2D-3 in Table V) with other approaches only relying on time representations of the signal, we repeated the experiments with two 1-dimensional CNN variants. The first one (CNN 1D-3), keeps the layer structure of the 2D CNN proposed (see Fig. 2) but with 1D temporal windows as inputs. The second one (CNN 1D-6) presents an increased complexity by duplicating the number of convolutional layers in CNN 1D-3. The comparative results are also shown in Table V, where a higher performance of the 2D CNN (CNN 2D-3 in the table) can be observed for all figures of merit.

To gain additional insight into the classification regions defined by the binary two-dimensional classifier, we show in Figs. 8 through 10 scatter diagrams along the 10 folds for the short through long activity terms. The scatter diagrams represent the percentage of windows classified as NON-ADHD for day

[1] Sensitivity=TP/(TP+FN), Specificity=TN/(TN+FP), Accuracy=(TP+TN)/(TP+TN+FP+FN), for AUC see [46].

[2] PPV=TP/(TP+FP), NPV=TN/(TN+FN), LR+=TP/FP, LR-=TN/FN [47].

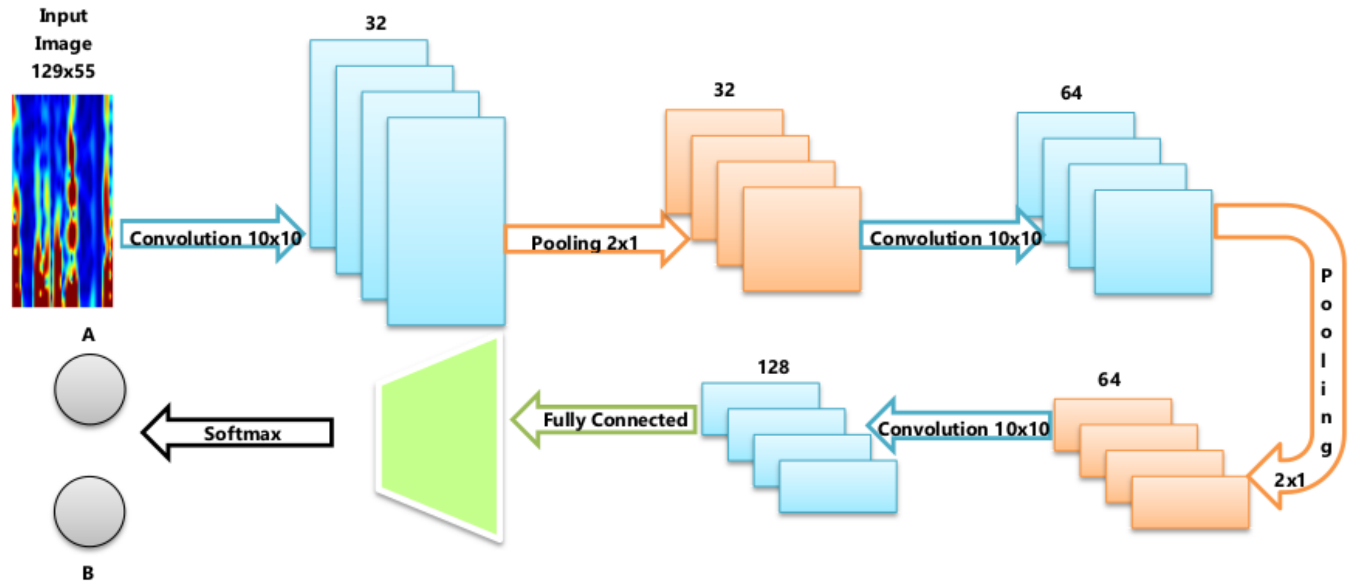

Fig. 2: Structure of the Convolutional network used for training

TABLE V: Results obtained for the global classifier ($70-30\%$, train/test) for the proposed 2D CNN with inputs in the time-frequency domain (CNN 2D-3). A comparison is provided with two 1D CNNs in the time domain: CNN 1D-3 has the same number of convolutional layers (3); CNN 1D-6 has 6 convolutional layers. Results show the mean and standard deviation (mean $\pm$ std) of the 10 folds described in Section III-B.3. For LR+, as there is a significant number of folds with 100 % specificity (yielding infinite LR+), minimum values are presented as a worst case scenario. Best results are boldfaced.

Window Size	CNN Type	Acc.	Sens.	Spec.	AUC
1800 s	CNN 2D-3	**0.9643± 0.0302**	**0.9429 ± 0.0514**	**0.9857± 0.023**	**0.9980± 0.029**
	CNN 1D-3	0.5571± 0.0999	0.3762 ± 0.2158	0.7381± 0.1128	0.5755± 0.1359
	CNN 1D-6	0.5595± 0.0834	0.5143 ± 0.203	0.6048± 0.2177	0.6002± 0.1157
300 s	CNN 2D-3	**0.9857 ± 0.0166**	**0.9762± 0.0337**	**0.9952± 0.0151**	**0.9993± 0.0022**
	CNN 1D-3	0.8476 ± 0.1072	0.8619± 0.1132	0.8333± 0.1669	0.9184± 0.1006
	CNN 1D-6	0.9286 ± 0.0745	0.9286± 0.1476	0.9286± 0.1059	0.9971± 0.0041
60 s	CNN 2D-3	**0.9691 ± 0.0252**	**0.9524± 0.0502**	**0.9857± 0.023**	**0.9918 ± 0.0117**
	CNN 1D-3	0.8476 ± 0.1072	0.8619± 0.1132	0.8333± 0.1669	0.8317 ± 0.2001
	CNN 1D-6	0.9167 ± 0.0393	0.9048± 0.055	0.9286± 0.0683	0.9619 ± 0.0321
Window Size	**CNN Type**	**PPV**	**NPV**	**LR+**	**LR-**
1800 s	CNN 2D-3	**0.9854± 0.0235**	**0.9474± 0.0483**	**18**	**0.0580 ± 0.055**
	CNN 1D-3	0.5727± 0.1666	0.5532± 0.0825	0.4286	0.8451 ± 0.2811
	CNN 1D-6	0.5834 ± 0.1253	0.5596 ± 0.0851	1.3012	0.8030 ± 0.2563
300 s	CNN 2D-3	**0.9955± 0.0144**	**0.9777± 0.0312**	**21**	**0.0238 ± 0.033**
	CNN 1D-3	0.8532 ± 0.1293	0.8614 ± 0.1101	1.5455	0.1657 ± 0.1571
	CNN 1D-6	0.9405 ± 0.0836	0.9456 ± 0.101	13	0.0768 ± 0.1476
60 s	CNN 2D-3	**0.9859± 0.0227**	**0.9560 ± 0.0448**	**20**	**0.0483 ± 0.050**
	CNN 1D-3	0.8923 ± 0.1254	0.7905 ± 0.144	1.3548	0.2152 ± 0.1966
	CNN 1D-6	0.9311 ± 0.0596	0.9094 ± 0.0482	12.667	0.1025 ± 0.0577

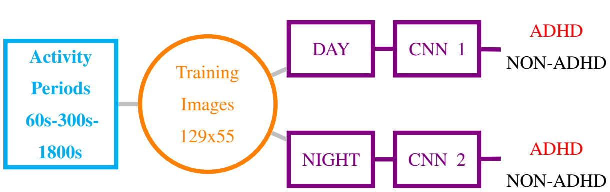

Fig. 3: Training process of the diagnostic system

(x–axis) and night (y–axis) periods, so cases are expected to show smaller values than controls. These were then used as inputs to the final SVM for patient diagnosis (note that the % of ADHD windows could have been used indistinctively as it can be directly obtained from the % of NON-ADHD windows, and the point clouds would have exchanged positions). Colors in the samples indicate whether the classification was correct or incorrect for that specific patient. The number of misclassified patients is significantly low, especially when 300s windows are employed (see Fig. 9). Specificity in this case is particularly high, with only one false positive (control identified as ADHD). The number of false negatives is also low (only 5 non-diagnosed patients).

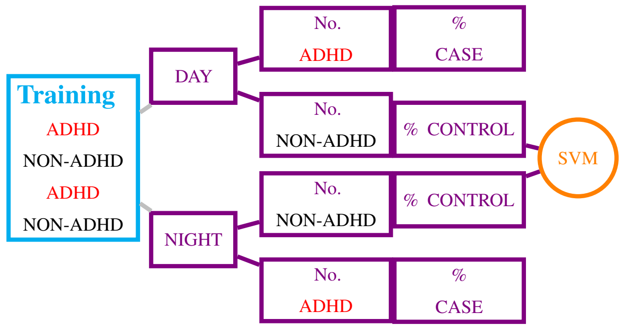

Fig. 4: Training Process for global classification

C. Classifier for small training set

When it comes to training a CNN, one of the factors to take into account is network robustness; a robust system should be able to find patterns in the training set even though it is smaller than the test set. To evaluate how the network behaves under these circumstances, table VI shows performance results for 10 folds with a splitting of 20% - 80% respectively for training and testing.

For this train/test proportion we also show the corresponding scatter diagrams associated to the 10 folds (Figs. 11 through 13 for short through long terms, respectively) . Clearly, the figures show a higher overlap degree with respect to the previous cases, but a high degree of separability between the two point clouds still remains.

V. Discussion

Results in Section IV suggest that actigraphy is a useful tool for ADHD diagnosis. Specifically, table V reveals that the cascade of a CNN and a SVM classifier gives rise to high classification figures for this application domain, also providing strong evidence from a medical point of view [47]. Medium-term activity windows have provided better results than the other two, and this leads us to interpret that medium term activities may be more versatile to find discriminative patterns for this problem. In addition, the results obtained using a training set of 20% of the whole dataset instead of 70%, reflects that the system is robust and efficient. In terms of accuracy we have reached figures on and above 0.90 for the 20% training case, as table VI shows; this loss of accuracy is approximately 8.59%, which does not seem excessive despite the small data fraction used for training; the values of LR+/LR- still provide high evidence on the suitability of the test.

Regarding the decision criterion, Figs. 8 and 13 provide interesting insight on how the classifier works; specifically, the boundary seems more directly related to the percentage value of day-time activity as opposed to that at night time. Clearly both activity percentages play a role, so a two dimensional classifier will work better than its one dimensional counterpart but,as indicated, day time activity seems a more discriminative dimension.

In Section II we have enumerated different studies that have faced the problem of ADHD diagnosis. Specifically, in [17] a classical pattern recognition approach was followed, obtaining **0.8571** accuracy, **0.9500** sensitivity, and **0.7727** a specificity. According to our results, we can state that the method here proposed has better figures, getting an accuracy, to sum up, equal to **0.9857**. These figures are obtained, in addition, with a larger sample, so our results are better as well as more robust than those we reported a few years ago.

Muñoz et al. report in their study [21] the use of CNN with actigraphs; however, shorter exam times have been tried, namely, only 6 school hours have been analyzed. In the case of Mahony et al [32] inertial measurement units (IMUs) with a SVM are used (details of both studies are summarized in table III). Comparing with our method, it should be noted that our study cohort is larger than the one used in both; in addition, our approach does not pose any restriction in children activity since our measures are taken in an average school day with routine activity while in [32] the authors propose a laboratory experiment. Moreover, we have obtained better accuracy figures than those reported in these references, as pointed out in Table V (recall from section II that accuracies reported in those references were 0.9375 and 0.9512 respectively); we stress that we do not need any sort of controlled environment and/or specifically designed experiment. We therefore avoid any bias caused by the experiment itself or, to say the least, biases are greatly diminished with our measurement procedure.

One of the main characteristics of using deep learning for any application and, in particular, for classification of spectrograms, is the implicit feature extraction and selection procure accomplished during the training stage; this has the direct consequence that the classifier is not limited to the features used as input by the designer —which are, in turn, selected after a somewhat tedious procedure [17]— but the classifier extracts and selects features on the fly, leading to an overall better performance since the mentioned limitation is avoided. The other side of the coin is the clinical interpretation of the activation of the network layers, which is not at all obvious. This is a path that should be explored in future works. In addition, the comparison in Table V has revealed that 2D processing in the Fourier domain yields better results than performing 1D processing directly in the signal domain for a similar CNN architecture, and even for a more complex one doubling the number of convolutional layers. The number of trainable parameters in this latter 1D CNN was approximately 50% that of the proposed spectrogram-based 2D CNN, but training such network required a significantly higher amount of time (approximately 4x). It may be arguable that a deeper 1D CNN network may reach comparable results in the signal domain as long as the number of free parameters is comparable as well. However, this approach does not seem worth taking considering the required training time, on one hand, and the low computational cost of periodogram calculation, on the other.

Regarding other data acquisition techniques, MRI and EEG have traditionally been used in the search for an objective diagnosis procedure of ADHD. Recent studies that merge the use of these techniques and SVM classifiers have obtained interesting results. These methods, however, are accompanied with a high complexity to obtain the data, not to mention their associated costs. Our approach, on the other side, obtained high performance figures with an affordable wristwatch. We

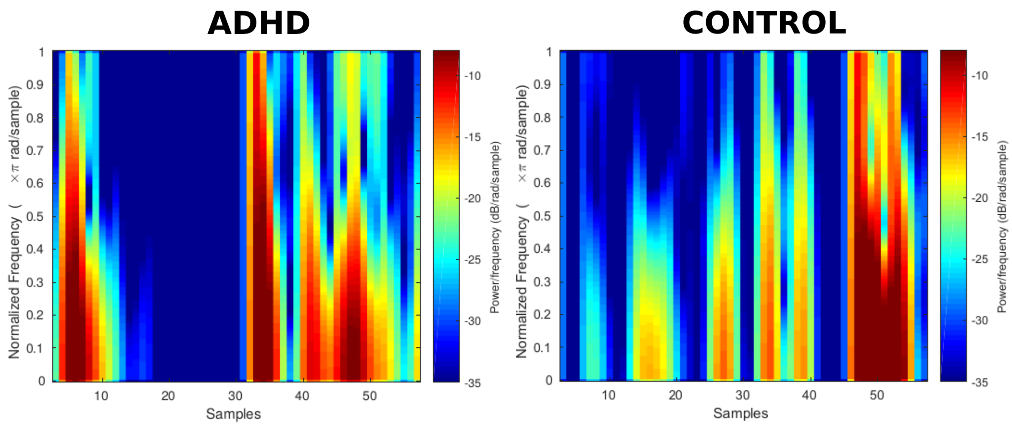

Fig. 5: Spectrograms obtained for an ADHD patient and a control (60s window).

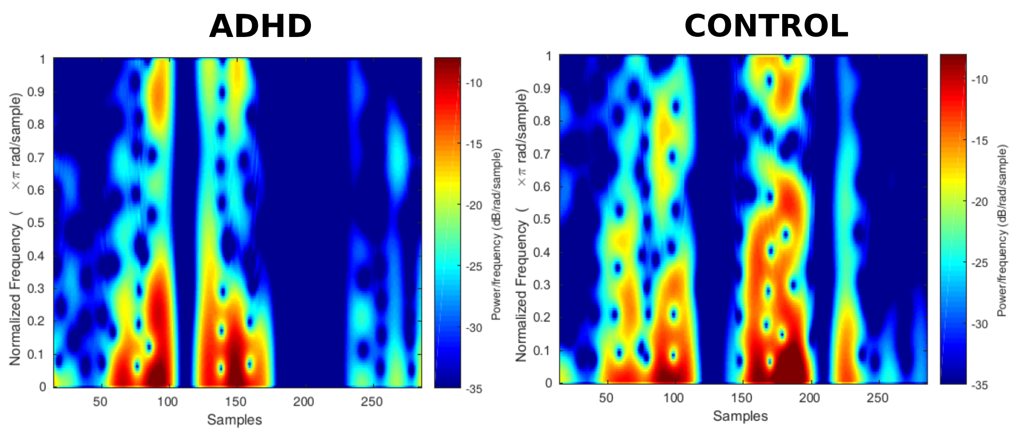

Fig. 6: Spectrograms obtained for an ADHD patient and a control (300s window).

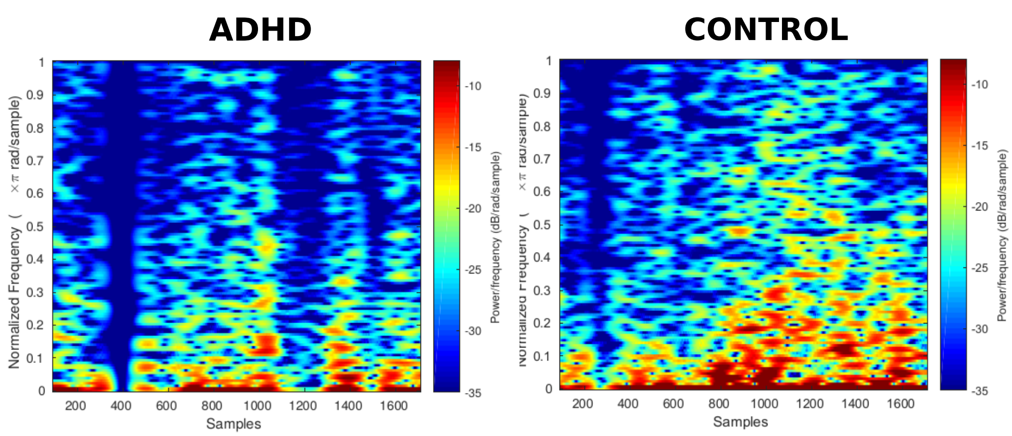

Fig. 7: Spectrograms obtained for an ADHD patient and a control (1800s window).

TABLE VI: Results obtained for the global classifier ($20-80\%$, train/test). Results show the mean and standard deviation (mean ± std) of the 10 folds described in Section III-B.3. For LR+, as there is a significant number of folds with 100 % specificity (yielding infinite LR+), minimum values are presented as a worst case scenario.

Window Size	**Acc.**	**Sens.**	**Spec.**	**AUC**
1800 s	0.8784±0.045	0.8172±0.119	0.9397±0.0447	0.9755±0.0087
300 s	0.9103±0.0216	0.8552±0.0502	0.9655±0.023	0.9837±0.0066
60 s	0.9078±0.0325	0.8776±0.0368	0.9379±0.057	0.9696±0.017
Window Size	**PPV**	**NPV**	**LR+**	**LR-**
1800 s	0.9371± 0.0426	0.8464 ±0.0762	7.57	0.1908 ± 0.1163
300 s	0.9623± 0.0228	0.8715±0.0395	5.10	0.1495 ± .0506
60 s	0.9371± 0.0554	0.8852 ±0.0291	10.40	0.1308 ± 0.0377

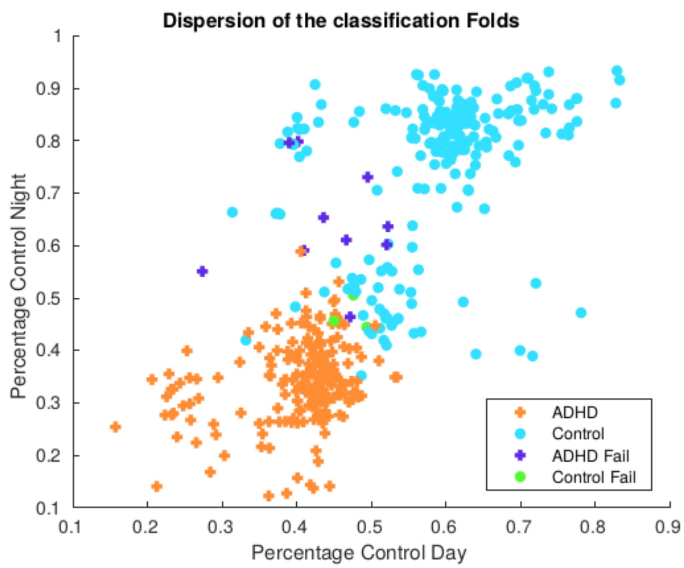

Fig. 8: Scatter diagram of the day vs. night percentage of NON-ADHD windows obtained from the CNN output (60s window, 70%-30% train-test split).

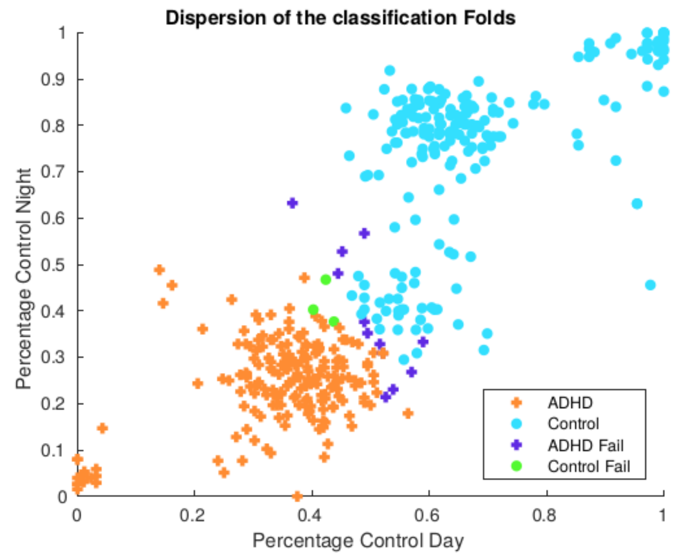

Fig. 10: Scatter diagram of the day vs. night percentage of NON-ADHD windows obtained from the CNN output (1800s window, 70%-30% train-test split).

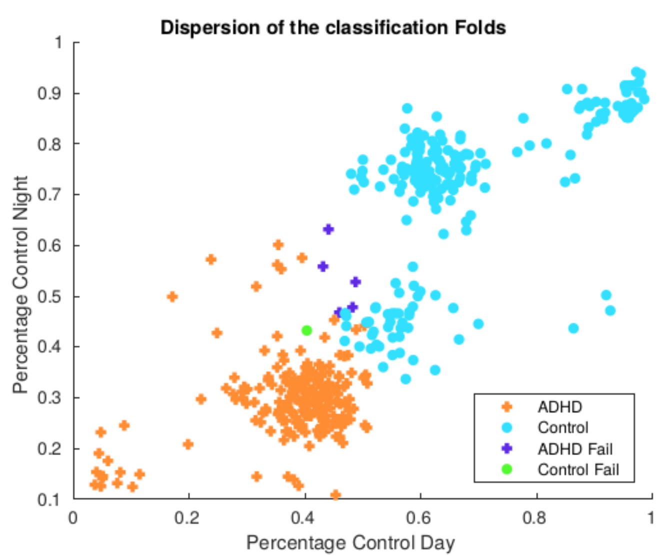

Fig. 9: Scatter diagram of the day vs. night percentage of NON-ADHD windows obtained from the CNN output (300s window, 70%-30% train-test split).

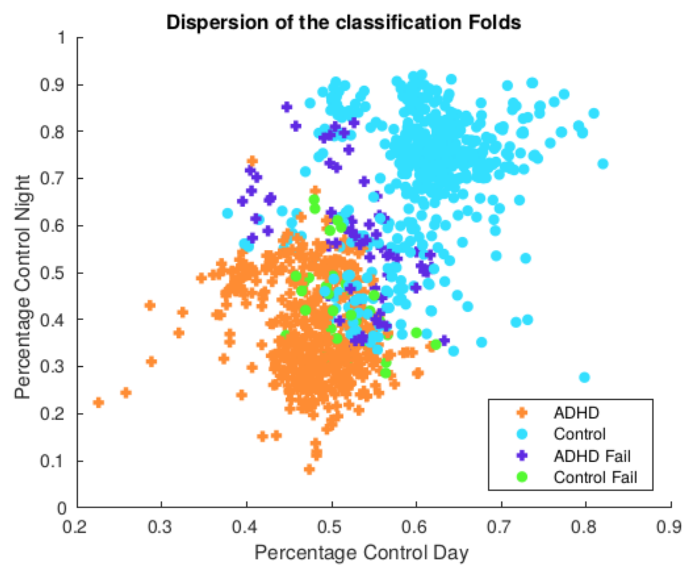

Fig. 11: Scatter diagram of the day vs. night percentage of NON-ADHD windows obtained from the CNN output (60s window, 20%-80% train-test split).

therefore consider that our method is a competitive candidate that could be used complementary to those methods, maybe as a useful and simple screening tool.

Despite the satisfactory results presented in this study, limitations in terms of the children study group must be taken into account. Our analysis has been done in a somewhat heterogeneous children sample, with a large age span and with no gender splitting (this was also the case in [17]). This is the reason why we have provided three different methodologies for diagnosing ADHD (according to the activity term) as opposed to a single decision rule. While the three methods are strongly supported by evidence [47], future research may reveal whether any of the three is more appropriate for any of the subgroups that may result from making the sample more homogeneous in terms of specific pathology, age and gender.

VI. Conclusion

This work reinforces the idea that combining deep learning techniques together with actimetry makes it possible to create a robust and efficient system for the objective ADHD diagnosis. Our results show comparable or even higher performance figures (accuracy, sensitivity and specificity for a case/control study) than other traditional methods, even though these methods make use of targeted experiments in a controlled environment. Additional parameters frequently used in the evaluation of diagnostic methods show strong evidence of the appropriateness of our methods. As previously discussed,

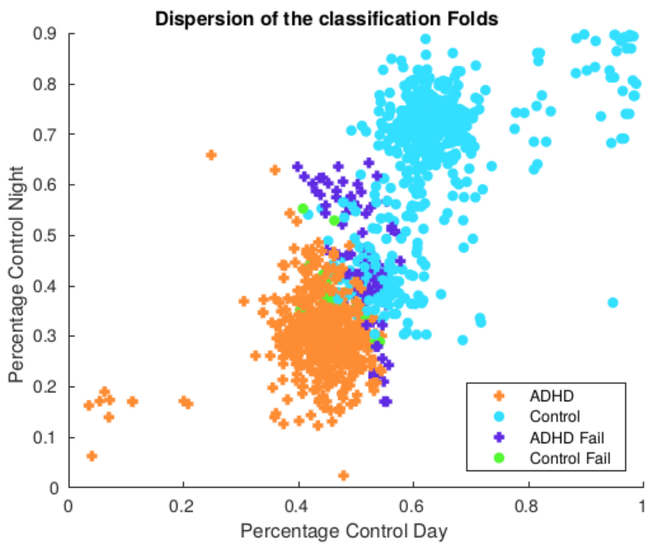

Fig. 12: Scatter diagram of the day vs. night percentage of NON-ADHD windows obtained from the CNN output (300s window, 20%-80% train-test split).

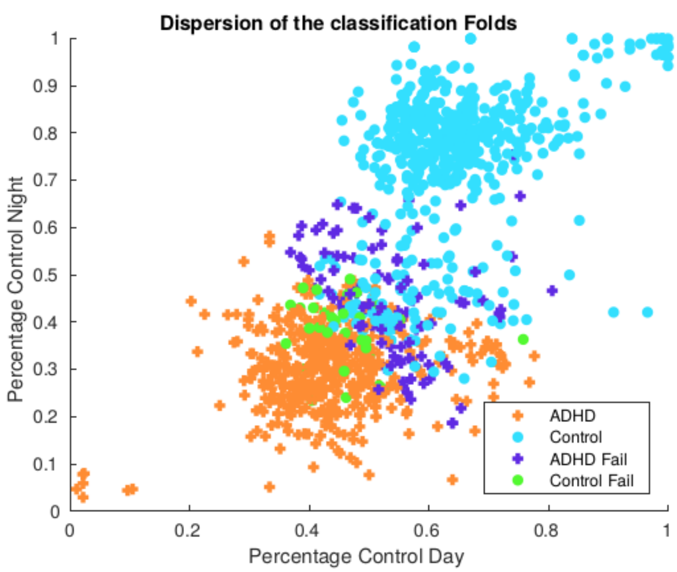

Fig. 13: Scatter diagram of the day vs. night percentage of NON-ADHD windows obtained from the CNN output (1800s window, 20%-80% train-test split).

finding clinical meaning to the features identified by the network is a matter of utmost importance, which requires further research.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

[1] L. Adler, T. Spencer, M. Stein, and J. Newcorn, "Best Practices in Adult ADHD: Epidemioólogy, Impairments and Differential Diagnosis," *CNS Spectr.*, vol. 13, pp. 4–5, 2008.

[2] R. Thomas, S. Sanders, J. Doust, E. Beller, and P. Glasziou, "Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis," *Pediatrics*, vol. 135, no. 4, pp. e994–e1001, 2015.

[3] F. Catalá-López, S. Peiró, M. Ridao, G. Sanfélix-Gimeno, R. Génova-Maleras, and M. A. Catalá, "Prevalence of attention deficit hyperactivity disorder among children and adolescents in spain: a systematic review and meta-analysis of epidemiological studies," *BMC psychiatry*, vol. 12, no. 1, p. 168, 2012.

[4] A. P. Association, "Diagnostic And Statistical Manual Of Mental Disorders, (DSM-5)," *American Journal of Psychiatry*, vol. 5ª edicion, 2013.

[5] J. López-Villalobos, J. Andrés-De Llano, M. López-Sánchez, L. Rodríguez-Molinero, M. Garrido-Redondo, A. Sacristán-Martín, and S. Martínez-Ribera, MT y Alberola-López, "Criterion validity and clinical usefulness of attention deficit hyperactivity disorder rating scale iv in attention deficit hyperactivity disorder (adhd) as a function of method and age," *Psicothema*, vol. 29, no. 1, pp. 103–110, 2017.

[6] L. Igual, J. C. Soliva, A. Hernández-Vela, S. Escalera, X. Jiménez, O. Vilarroya, and P. Radeva, "A fully-automatic caudate nucleus segmentation of brain mri: Application in volumetric analysis of pediatric attention-deficit/hyperactivity disorder," *BioMedical Engineering OnLine*, vol. 10, no. 1, p. 105, Dec 2011.

[7] S. Cortese, R. Azoulay, F. Castellanos, F. Chalard, M. Lecendreux, D. Chechin, R. Delorme, G. Sebag, A. Sbarbati, M.-C. Mouren, B. Dalla Bernardina, and E. Konofal, "Brain iron levels in attention-deficit/hyperactivity disorder: A pilot MRI study. the world journal of biological psychiatry," *The official journal of the World Federation of Societies of Biological Psychiatry.*, vol. 13, pp. 223–31, 2011.

[8] D. Yu, "Additional brain functional network in adults with attention-deficit/hyperactivity disorder: A phase synchrony analysis," *PLOS ONE*, vol. 8, 01 2013.

[9] A. Guzmán, E. Pereda, S. Mañas, L. D. Méndez, A. González, and J. J. González, "Electroencephalography signatures of attention-deficit/hyperactivity disorder: clinical utility," *Clinical Neurophysiology*, 2015.

[10] M. Ahmadlou, R. Rostami, and V. Sadeghi, "Which attention-deficit/hyperactivity disorder children will be improved through neurofeedback therapy? A graph theoretical approach to neocortex neuronal network of ADHD," *Neuroscience Letters*, vol. 516, no. 1, pp. 156–160, Dec 2012.

[11] T. Liu, Y. Chen, P. Lin, and J. Wang, "Small-World Brain Functional Networks in Children With Attention-Deficit/Hyperactivity Disorder Revealed by EEG Synchrony," *Clinical EEG and neuroscience*, 2014.

[12] H. Öztoprak, M. Toycan, Y. K. Alp, O. Arıkan, E. Doğutepe, and S. Karakaş, "Machine-based classification of ADHD and non-ADHD participants using time/frequency features of event-related neuroelectric activity," *Clinical Neurophysiology*, vol. 128, no. 12, pp. 2400–2410, 2017.

[13] A. Lenartowicz and S. K. Loo, "Use of EEG to diagnose ADHD," *Current Psychiatry Reports*, vol. 16, no. 11, p. 498, Sep 2014.

[14] A. S. Walters, R. Silvestri, M. Zucconi, R. Chandrashekariah, and E. Konofal, "Review of the possible relationship and hypothetical links between attention deficit hyperactivity disorder (ADHD) and the simple sleep related movement disorders, parasomnias, hypersomnias, and circadian rhythm disorders," *Journal of Clinical Sleep Medicine*, vol. 4, no. 6, pp. 591–600, 2008.

[15] A. Chesson, R. Ferber, J. Fry, M. Grigg-Damberger, K. Hartse, T. Hurwitz, S. Johnson, M. Littner, G. Kader, G. Rosen, R. Sangai, W. Schmidt-Nowara, and A. Sher, "Practice Parameters for the Indications for Polysomnography and Related Procedures," *Sleep*, vol. 20, pp. 406 – 422, 1997.

[16] A. Hvolby, "Associations of sleep disturbance with ADHD: implications for treatment," *ADHD Attention Deficit and Hyperactivity Disorders*, vol. 7, no. 1, pp. 1–18, Mar 2015.

[17] D. Martín Martínez, P. Casaseca de la Higuera, S. Alberola López, J. Andrés de Llano, J. López Villalobos, J. Ardura Fernández, and C. Alberola López, "Nonlinear Analysis Of Actigraphic Signals For The Assessment Of The Attention-Deficit/Hyperactivity Disorder (ADHD)," *Medical Engineering & Physics*, vol. 34, no. 9, pp. 1317–1329, 2012.

[18] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, T. Arora, and S. Taheri, "Sleep quality prediction from wearable data using deep learning," *JMIR mHealth and uHealth*, vol. 4, no. 4, 2016.

[19] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, S. Taheri, and T. Arora, "Impact of physical activity on sleep: a deep learning based exploration," *arXiv preprint arXiv:1607.07034*, 2016.

[20] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel eeg using convolutional neural networks," *arXiv preprint arXiv:1610.01683*, 2016.

[21] M. Muñoz-Organero, L. Powell, B. Heller, V. Harpin, and J. Parker, "Automatic extraction and detection of characteristic movement patterns in children with ADHD based on a convolutional neural network (CNN) and acceleration images," *Sensors*, vol. 18, no. 11, p. 3924, 2018.

[22] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, "Fusion of fMRI and non-imaging data for ADHD classification," *Computerized Medical Imaging and Graphics*, vol. 65, pp. 115–128, 2018.

[23] H. Sun, Y. Chen, Q. Huang, S. Lui, X. Huang, Y. Shi, X. Xu, J. A. Sweeney, and Q. Gong, "Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: A radiomics analysis," *Radiology*, vol. 287, no. 2, pp. 620–630, May 2018.

[24] S. Holley, C. Hill, and J. Stevenson, "An Hour Less Sleep is a Risk Factor for Childhood Conduct Problems," *Child Care Hlth. Dev.*, vol. 37, pp. 563–570, 2011.

[25] M. Dabkowska, T. Pracki, and D. Pracka, "The Objective Measurement of Movement vs the Intensification of ADHD Symptoms in Assessment of Parents and Doctors," *Medical and Biological Sciences*, vol. 26, pp. 11–17, 2012.

[26] M. Snitselaar, M. Smits, K. van der Heijden, A. Smit, and J. Spijker, "Influence of Methylphenidate on Circadian Rhythmicity and Sleep in Adult Attention–Deficit/Hyperactivity Disorder," *Sleep Biol. Rhythms*, vol. 11, pp. 282–285, 2013.

[27] P. Casaseca-de-la-Higuera, D. Martín-Martínez, S. Alberola-López, J. María Andrés-de-Llano, J. Antonio López-Villalobos, J. Ramón-Garmendia Leiza, and C. Alberola-López, "Automatic diagnosis of adhd based on multichannel nonlinear analysis of actimetry registries," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2012, pp. 4204–4207.

[28] S. Wiebe, J. Carrier, S. Frenette, and R. Gruber, "Sleep and Sleepiness in Children with Attention–Deficit/Hyperactivity Disorder and Controls," *J. Sleep Res.*, vol. 22, pp. 41–49, 2013.

[29] K. Gamble, R. May, R. Besing, A. Tankersly, and R. Fargason, "Delayed Sleep Timing in Adults with Attention–Deficit/Hyperactivity Disorder: A Controlled Actigrapy Study," *Chronobiol. Int.*, vol. 30, pp. 598–606, 2013.

[30] V. Moreau, N. Rouleau, and C. Morin, "Sleep of Children with Attention Deficit Hyperactivity Disorder: Actigraphy and Parental Reports," *Behav. Sleep Med.*, vol. 12, pp. 69–83, 2013.

[31] K. L. Hudec, R. M. Alderson, L. J. Kasper, and C. H. G. Patros, "Working memory contributes to elevated motor activity in adults with adhd: An examination of the role of central executive and storage/rehearsal processes," *Journal of Attention Disorders*, vol. 18, no. 4, pp. 357–368, 2014.

[32] N. O'Mahony, B. Florentino-Liano, J. J. Carballo, E. Baca-García, and A. A. Rodríguez, "Objective diagnosis of ADHD using IMUs," *Medical engineering & physics*, vol. 36, no. 7, pp. 922–926, 2014.

[33] R. Alderson, M. Rapport, L. Kasper, D. Sarver, and M. Kofler, "Hyperactivity in Boys with Attention Deficit/Hyperactivity Disorder (ADHD): The Association Between Deficient Behavioral Inhibition, Attentional Processes, and Objectively Measured Activity," *Child Neurpsychol.*, vol. 18, pp. 487–505, 2012.

[34] M. Nakatami, S. Okada, S. Shimizu, I. Mohri, Y. Ohno, M. Taniike, and M. Makikawa, "Body Movement Analysis During Sleep for Children with ADHD Using Video Image Processing," in *Proceedings of the International Annual Conference of the IEEE-Engineering in Medicine and Biology Society, 2013. EMBC.*, 2013, pp. 6389–6392.

[35] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, "The Neuro Bureau ADHD-200 Preprocessed repository," *NeuroImage*, vol. 144, pp. 275 – 286, 2017, data Sharing Part II. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105381191630283X

[36] ActiGraph wGT3X-BT. (Last access June 2019) https://www.actigraphcorp.com/actigraph-wgt3x-bt/.

[37] ActiGraph White Paper:What is a Count? (Last access June 2019) https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2017/11/26205758/ActiGraph-White-Paper_What-is-a-Count_.pdf.

[38] D. Martín Martínez, P. Casaseca de la Higuera, J. Andrés de Llano, J. Garmendia Leiza, S. Alberola-López, and C. Alberola López, "Automatic Detection Of Wakefulness And Rest Intervals In Actigraphic Signals: A Data-Driven Approach," *Medical Engineering & Physics*, vol. 36, no. 12, pp. 1585–1592, 2014.

[39] D. Wagner, K. Kalischewski, J. Velten, and A. Kummert, "Activity recognition using inertial sensors and a 2-d convolutional neural network," in *Multidimensional (nD) Systems (nDS), 2017 10th International Workshop on*. IEEE, 2017, pp. 1–6.

[40] MathWorks MATLAB for Deep Learning. (Last Access December 2018) `https://es.mathworks.com/`.

[41] A. Jordao, L. A. B. Torres, and W. R. Schwartz, "Novel approaches to human activity recognition based on accelerometer data," *Signal, Image and Video Processing*, vol. 12, no. 7, pp. 1387–1394, 2018.

[42] L. Granovsky, G. Shalev, N. Yacovzada, Y. Frank, and S. Fine, "Actigraphy-based sleep/wake pattern detection using convolutional neural networks," *arXiv preprint arXiv:1802.07945*, 2018.

[43] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, pp. 1488–1492.

[44] Y. Lang, C. Hou, Y. Yang, D. Huang, and Y. He, "Convolutional neural network for human micro-doppler classification," in *Proc. European Microwave Conference*, 2017.

[45] S. Y. Kung, *Support vector machines*. Cambridge University Press, 2014.

[46] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[47] J. Deeks and D. Altman, "Diagnostic tests 4: likelihood ratios," *British Medical Journal*, vol. 329, pp. 168–169, 2004.