



OPEN

Human activity recognition of children with wearable devices using LightGBM machine learning

Gábor Csizmadia^{1✉}, Krisztina Liszkai-Peres^{1,3,4}, Bence Ferdinandy², Ádám Miklósi^{1,2} & Veronika Konok¹

Human activity recognition (HAR) using machine learning (ML) methods has been a continuously developed method for collecting and analyzing large amounts of human behavioral data using special wearable sensors in the past decade. Our main goal was to find a reliable method that could automatically detect various playful and daily routine activities in children. We defined 40 activities for ML recognition, and we collected activity motion data by means of wearable smartwatches with a special SensKid software. We analyzed the data of 34 children (19 girls, 15 boys; age range: 6.59–8.38; median age = 7.47). All children were typically developing first graders from three elementary schools. The activity recognition was a binary classification task which was evaluated with a Light Gradient Boosted Machine (LGBM) learning algorithm, a decision tree based method with a threefold cross validation. We used the sliding window technique during the signal processing, and we aimed at finding the best window size for the analysis of each behavior element to achieve the most effective settings. Seventeen activities out of 40 were successfully recognized with AUC values above 0.8. The window size had no significant effect. In summary, the LGBM is a very promising solution for HAR. In line with previous findings, our results provide a firm basis for a more precise and effective recognition system that can make human behavioral analysis faster and more objective.

In behavioural sciences, the objective and quantifiable measurement of behavior is a fundamental requirement for conducting research. For this purpose, video-based or live behavior coding is a frequently used method. However, manual coding (i.e. coding by human observers based on visual inspection) is time-consuming and can have subjective components¹.

Body movements are among the main measurable components of behavior. Therefore, the use of motion sensor devices, such as accelerometers and gyroscopes, can help behavioral scientists to automatically measure behavior, both in studies on animals^{2,3} and humans^{4,5}. These devices can supply researchers with a large amount of objective and quantitative data based on the three spatial dimensions. This way, data collection can be conducted with more individuals in parallel, and over a longer period of time because these devices can be used in a wide range of environments (not only in the laboratory). This may help to solve statistical problems related to low sample size (and consequences e.g. low statistical power, not representative sample).

However, such data (big data) requires specific analytic methods. Machine learning algorithms can be used to train models that are able to automatically identify predetermined behavior categories (supervised learning⁶). This can improve behavior measurement and make it more objective than the visual inspection of behavior by human observers.

In humans, movement sensors are increasingly used in different settings, such as the entertaining or healthcare industry. Smartwatches with accelerometer and gyroscope can derive general health related parameters such as total step counts, and people often use this function for maintaining regular physical activity. These devices can be used also by healthcare professionals, for e.g. assisting rehabilitation^{7,8}, monitoring physical activity in specific patient groups^{9,10}, or monitoring symptoms of movement disorders (e.g. essential tremor or motor symptoms of Parkinson's disease; for a review see¹¹).

Sedentary lifestyle is increasingly frequent in children which is partly attributed to the excessive usage of digital devices¹². This may explain why the prevalence of obesity, diabetes and other related health problems is increasing in childhood^{13,14}. However, children can be especially motivated by games, and the ubiquity of smart

¹Department of Ethology, Eötvös Loránd University, Budapest, Hungary. ²MTA-ELTE Comparative Ethology Research Group, Budapest, Hungary. ³Doctoral School of Psychology, Eötvös Loránd University, Budapest, Hungary. ⁴Institute of Psychology, Eötvös Loránd University, Budapest, Hungary. ✉email: csizmadia.gabor.tamas@ttk.elte.hu

mobile devices makes it possible to use these devices for exergaming and facilitating children to do physical activity. An additional advantage of mobile devices is that they can be used anywhere, therefore, outdoor activities can be also facilitated by them, and geolocation data can be also exploited for the game experience¹⁵. If machine learning based feedback can be included in a game, it may become more motivating and entertaining and can also improve motor skills (for a review see¹⁶).

Another potential application area for motion sensors with machine learning in healthcare is the identification of symptoms of movement disorders or neurological diseases, which can aid either in the establishment of a diagnosis¹⁷ or in the management of symptoms^{18,19}. This method could help in diagnosing mental disorders which have no biomarker and thus, straightforward diagnosis is problematic. For example, activity-based automatic detection has been increasingly used in case of neurocognitive disorders, like autism spectrum disorder and attention-deficit hyperactivity disorder²⁰. These developmental disorders start in infancy but are often diagnosed only in pre-school or school years, while earlier diagnosis leads to better prognosis²¹. Therefore, the usage of motion sensor data with machine learning has the potential to predict developmental disorders at an earlier age than usual²². Additionally, this method can also aid in managing symptoms that present a burden to the affected individuals and their family^{23,24}.

In sum, wearable motion sensors can provide an objective research tool and a useful healthcare aid in prevention, diagnosis, and treatment. However, more research is needed on how automatic detection of activity can be optimally carried out. We propose a method with which children's activity can be assessed and detected automatically. This way game applications can be improved which facilitate physical activity, predict neurodevelopmental disorders and/or improve motor skills in children with motor problems.

We developed a wearable system and an activity test battery for children and assumed that we can successfully identify activities by the means of machine learning models. In contrast to other approaches⁴, we aimed at applying and detecting complex (playful and everyday) activities which can form the base of a game application.

As the time spent on using mobile devices, especially playing games, increases in the school age²⁵ and many children get diagnosis of developmental disorders at the beginning of elementary school²⁶, both the development of game applications facilitating physical activity and the training of machine learning models to predict developmental disorders would be feasible to carry out in this age group. Therefore, in the present study we focused on 6–8-years-old children.

ML methods in Human Activity Recognition (HAR) rely on two main approaches regarding the pre-processing of data before the learning task, one is based on derivative parameters or features, and the other one works on raw data. The first one requires the segmentation of the data and the extraction of derived features. On these data mostly decision tree-based classifications are used. The second one, is the deep learning method, which automatically learns the required feature representation directly from the raw data using e.g. convolutional neural networks²⁷.

The first step of any ML method is *segmentation*, i.e. dissecting the time series into smaller segments. One of three different windowing techniques are usually used to determine these segments and recognition accuracy may be affected by them and by the length of the window: (i) sliding window technique where signals are divided into equal segments using a sliding fix-length window; (ii) event-defined window techniques, where pre-processing is necessary to locate specific events, which are further used to define successive data partitioning and (iii) activity-defined window technique where data partitioning is based on the detection of activity changes. These windows may (but not necessarily) overlap, and the degree of the overlap may also affect the ML performance²⁸. The sliding window approach is well-suited to real-time applications since it does not require pre-processing.

The second step of many ML methods is the *feature extraction* the aim of which is to get the most effective features from the obtained raw segments. Time domain features include mean, median, variance, skewness, kurtosis, range etc. Peak frequency, peak power and spectral power on specific frequency bands and spectral entropy are generally included in the frequency-domain features. Feature extraction is crucial in any HAR targeted solution since the features used primarily determine the overall system accuracy. One widely used approach is to rely on various arbitrarily chosen measures of the raw data and then to find the most effective combination of these features (these may range from a few to many hundreds).

After segmentation and feature extraction, various classification algorithms are applied to each window, and the recognition process (decision tree based methods, most frequently boosted variants) runs on the derivative features.

There is no unified standpoint about which ML methods (including classification method, or preprocessing method) yield the best success. The choice of the classifier method is the most important parameter, followed by the segmentation method, window size and finally sampling frequency²⁹. The picture is even more complex because there is also a correlation between the number of features and the window size³⁰.

Therefore, our aim was to find the best machine learning method and parameters that delivers the best performance in the recognition process, is fast and effective and can also run on different types smartphones or smartwatches, so it is usable in typical life situations. Further, we aimed also to produce guidelines for other similar HAR projects targeting children's behavior.

Results

The overall (mean) accuracy of the classification was 0.95 ± 0.04 ($M \pm SD$) and the AUC was 0.76 ± 0.15 . The performance of the LGBM model regarding the recognition of a specific activity varied from AUC = 0.5 (Shoe_off_same, Sock_off_other) to AUC = 0.98 (Hopscotch) (Table 1).

As the underlying data is inherently imbalanced, the accuracy measures could be misleading, but we provided them only for comparison with similar data in the literature, so we favour AUC as the main performance

Activity name	Description	Sequence	Activity type	Accuracy	AUC	Num pos	Pos %
Hopscotch	Alternating hopping on single leg and double leg on a hopscotch ground	One jump (from bouncing off till arriving)	Playful	0.9773	0.9871	868	3.97
Ball	Throwing and catching a ball	From preparing to throw (hand rising) till catching the ball	Everyday	0.9813	0.9569	556	2.54
Goliath	Walking on the toes, hands stretched upwards	From the first step till the last step; if child stops coding is paused	Playful	0.9769	0.9447	765	3.50
Drawing	Drawing with a pencil on a paper	As the pencil touches the paper till lifting up the pencil	Playful	0.9266	0.9429	2656	12.15
Crab	Crawling backward (hands stretched, legs bent, chest upwards)	From the first step till the last step; if child stops coding is paused	Playful	0.9649	0.9396	880	4.03
Swimming	Lying on stomach, hands moving around like swimming	From the first hand movement till end of the last	Playful	0.9740	0.9380	1037	4.74
Spider	Crawling forward (hands stretched, legs bent, chest upwards)	From the first step till the last step; if child stops coding is paused	Playful	0.9573	0.9314	947	4.33
Seal	Legs and hips on the ground, hands stretched, moving only by using hands	From the first step till the last step; if child stops coding is paused	Playful	0.9461	0.9159	902	4.13
Building blocks	Building a tower from 5 building blocks (coding only when the action is done with the hand smartwatch on it, and only if building a horizontal tower)	From reaching towards a cube till putting it onto another cube	Everyday	0.9329	0.9057	1773	8.11
Bear	Crawling forward (hands stretched, legs bent, chest downwards)	From the first step till the last step; if child stops coding is paused	Playful	0.9588	0.9020	787	3.60
Light off	Turning off the light, then releasing hands	From lifting hand till the hand is next to the body in the start position	Everyday	0.9820	0.8984	359	1.64

Table 1. The Budapest activity test battery (BATB) and performance of the machine learning methods. The table was sorted by AUC with the highest recognition figures at the top. Num pos is the number of occurrences of a given activity in the dataset.

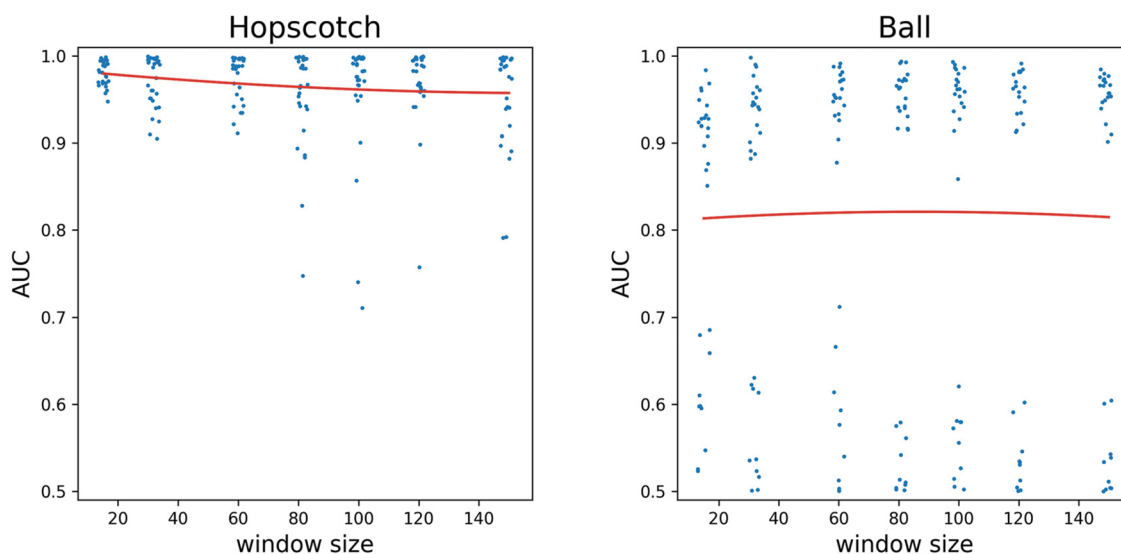


Figure 1. Playful activities with the highest AUC values, AUC values as a function of the window size. Quadratic curves fitted (Hopscotch: adj. $R^2 = 0.240$, Ball: adj. $R^2 = 0.044$). A small random jitter was added to the window sizes for better visibility.

indicator. The recognition was the highest ($AUC > 0.9$) in the case of Hopscotch, Ball, Goliath, Drawing, Crab, Swimming, Spider, Seal, Building blocks and Bear (Figs. 1, 2, 3).

To analyse whether there is an optimal window size in terms of AUC, we fitted a quadratic regression using the `lm` function of R, version 3.6.3, on the AUC—window size values of each activity. A quadratic curve allows for having a local maximum, enabling the investigations of the optimum. We defined the effect size as relevant if due to window size change an AUC change of 0.05 can be achieved within the 15–149 range explored. We assessed adj. R^2 values of the regression for goodness of fit. The highest value of R^2 was 0.327 for “Seal” (effect size = 0.027), while the highest effect size was 0.156 for Rabbit ($R^2 = 0.069$, see Fig. 2), with the effect size mostly due to the linear term in the model, that is, no local maxima was found. “Door handle” is the sole activity where both measures were appreciable ($R^2 = 0.092$, effect size = 0.062), while for most activities, both measures were

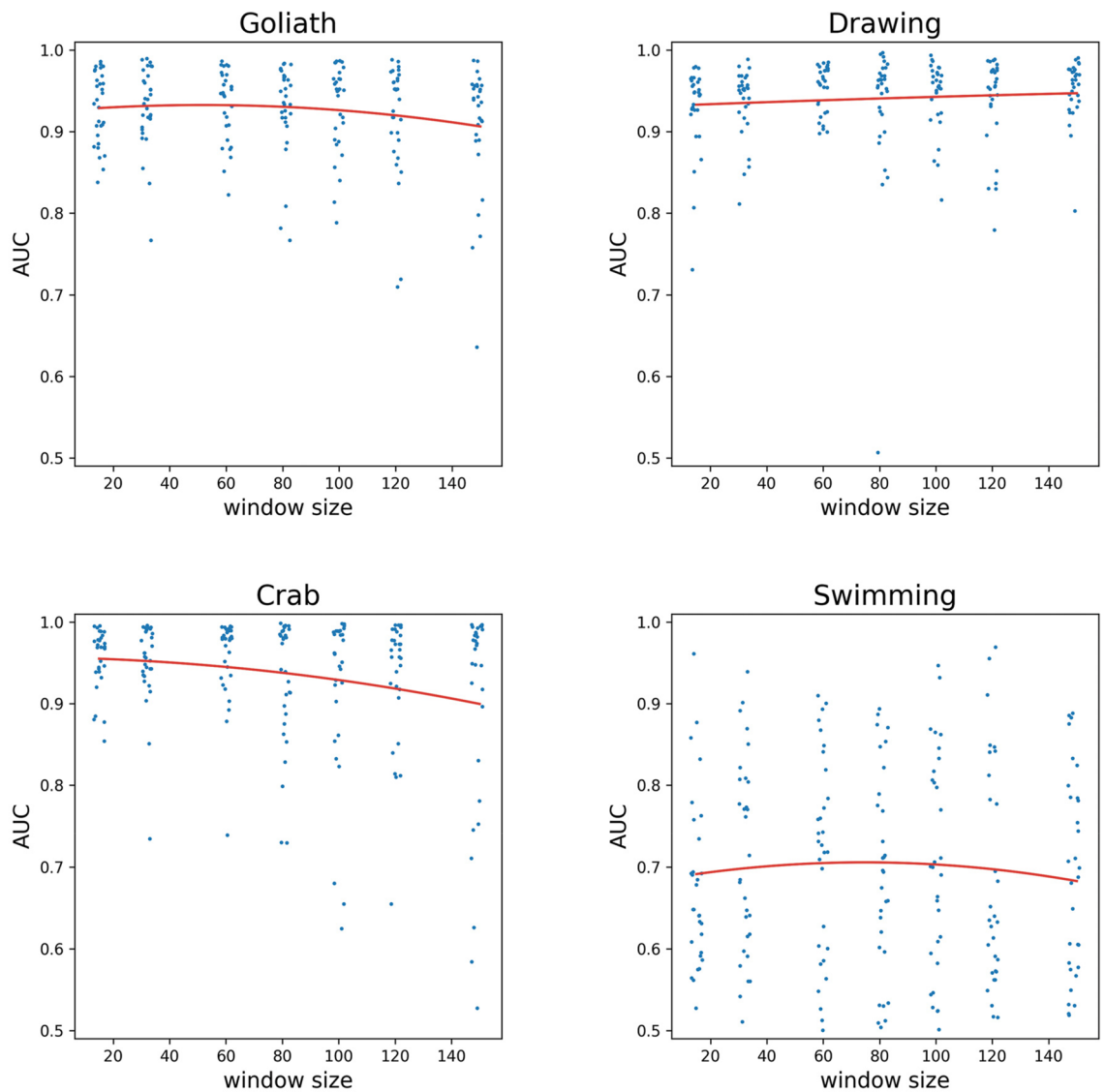


Figure 2. Playful activities with the highest AUC values, AUC values as a function of the window size. Quadratic curves fitted (Goliath: adj. $R^2=0.007$, Drawing: adj. $R^2=0.074$, Crab: adj. $R^2=0.084$, Swimming: adj. $R^2=0.026$). A small random jitter was added to the window sizes for better visibility.

low (see Fig. 4 and Tables 1, 2, 3). Thus, for all but one activity the applied window size did not influence the AUC values.

Discussion

Our main goal was to introduce a ML method which is able to recognize various playful and everyday activities for rapid automated analysis in children. Overall, our method of data pre-processing and the applied LGBM algorithm is successful regarding the recognition performance. Our mean accuracy (0.95) is in the top range of other similar HAR machine learning results (see summary Table 1 in³¹) When evaluating our results, it should be noted that we used only one movement sensor located on the wrist of the children.

Although the accuracy values were excellent, we also examined AUC values because those are more robust indicators of recognition success (e.g. accuracy is high even when not only true positives, but false positives are high). We obtained at least acceptable AUC values for all activities, and very good values for the majority of them, e.g. seventeen activities were recognized with AUC values above 0.8. Activities with lower AUC values are those with lower sample size (see Table 1, last column for the occurrence of the activities). Therefore, larger sample might increase the recognition rate of these activities.

We achieved the highest AUC values mainly for the playful activities. Besides that these were also among the activities most frequently carried out, but it is also possible that the LGBM method performs better in the case of complex activities that involve several body parts, with characterized sensor data patterns³².

The recognition of everyday activities was in the lower segment of the performance list. One possible explanation (besides lower sample size) is that some of these actions are executed very similarly in 3D space. For example,

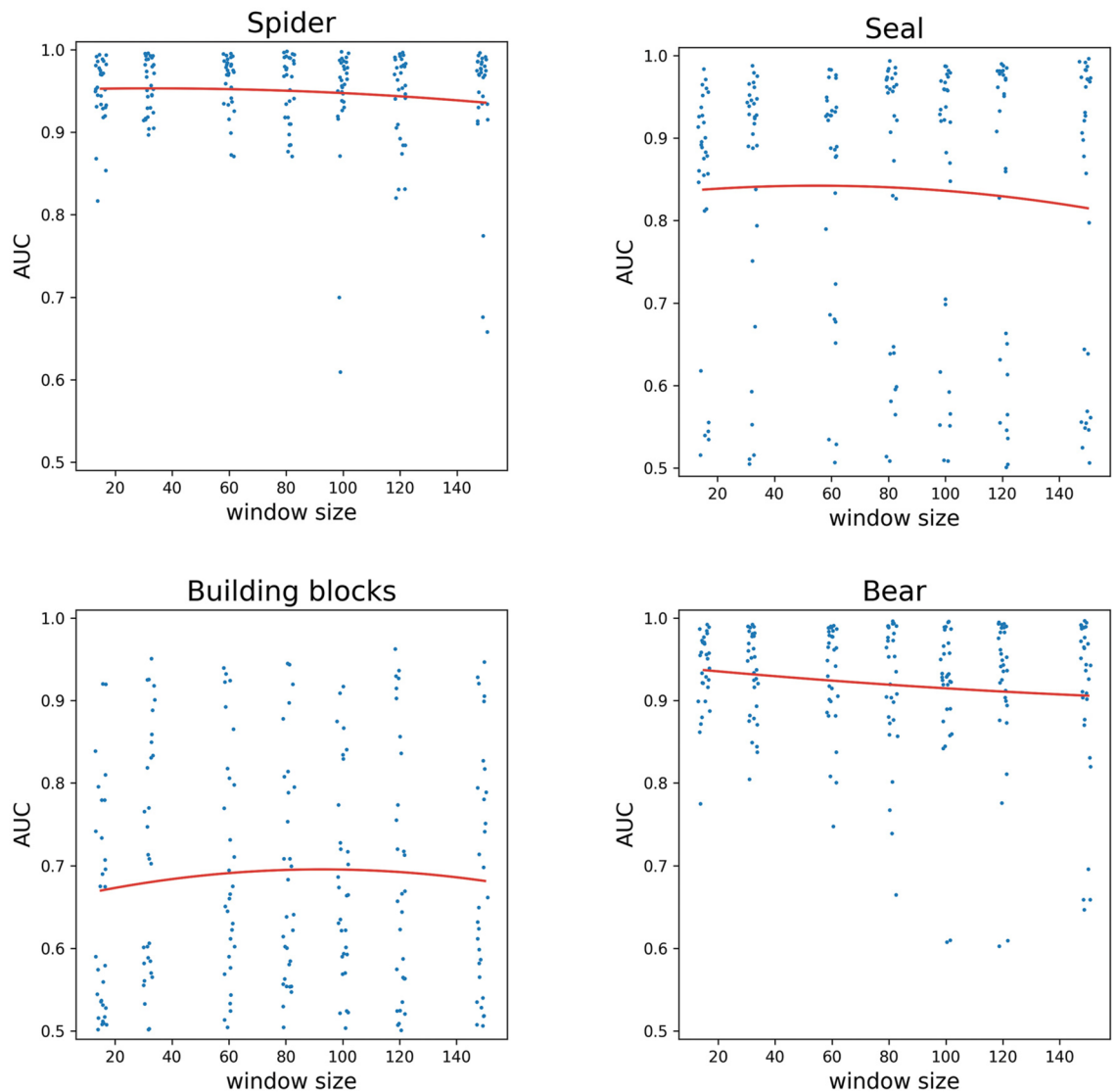


Figure 3. Playful activities with the highest AUC values, AUC values as a function of the window size. Quadratic curves fitted (Spider: adj. $R^2=0.091$, Seal: adj. $R^2=0.327$, Building blocks: adj. $R^2=0.036$, Bear: adj. $R^2=0.115$). A small random jitter was added to the window sizes for better visibility.

praying and clapping with the hands are very similar actions when measured by one sensor on the wrist that mainly differ in their speed of execution. This kind of problem, undeniably, could be addressed by using different feature extraction methods for various activities and optimising for unique features from the best performed IMU signals, but our intention was to develop a widely usable method which could be adopted for any kind of activity in the future. Another explanation could be that some everyday activities (e.g. taking socks on or off) could be executed variably resulting in large inter-individual differences. Such variation may make it impossible for the algorithm to find a common pattern in the feature set.

No systematic effect of the window size (window size 15 (0.3 s) to 149 (3 s)) on the AUC was revealed, suggesting that it either does not exist, or the sample size and range of values was insufficient to detect such an effect. Awais et al.³³ reported similar observations with no or little effect of the window size when they compared various experiments on HAR datasets. In contrast, Banos et al.³⁰ reported a significant drop in the accuracy when they examined the effect of the window size ranging from 0.25 s to 7 s in steps of 0.25 s. Intervals below 1 s resulted in lower accuracy values. Based on our data, window size optimization has an effect only when the recognition performance is in the middle range, not too low or not too high. Both scenarios push the variance toward extreme values, so the outcome is masked by this side effect. Importantly, we used subject-independent cross-validation (CV), while window size effect was reported in studies applying a subject-dependent CV³⁰. In the latter case individual differences do not mask the effect of window sizes. This is especially true for overlapping windows, which we used in this study, in case of which the performance difference between subject-dependent CV and subject independent CV increases with the window size. Dehgani et al.²⁸ also reported that there was no window size effect in case of subject-independent CV.

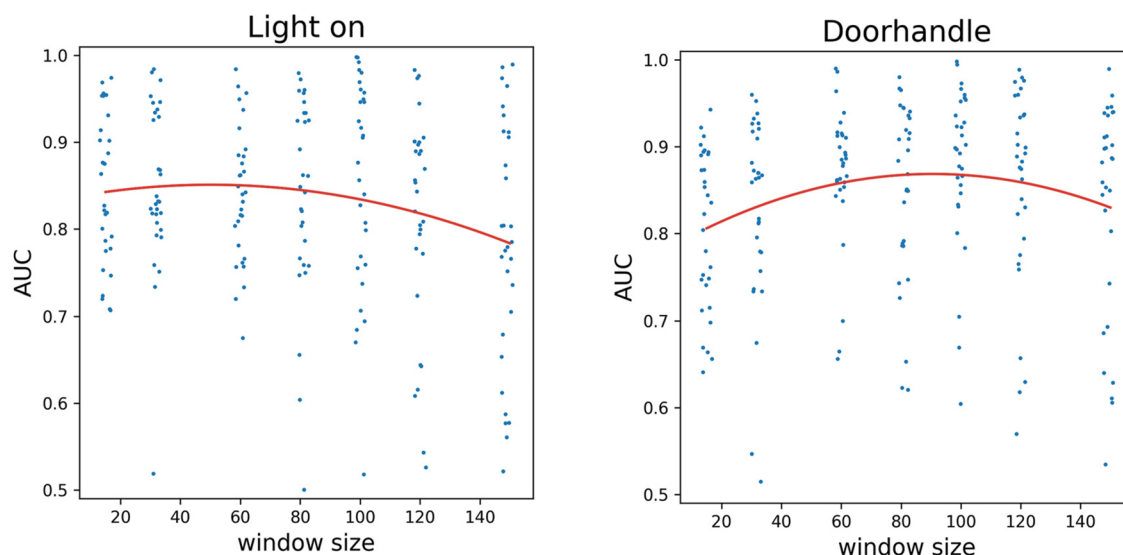


Figure 4. The AUC values of two everyday activities as a function of the window sizes and the quadratic curves fitted (Light on: adj. $R^2 = 0.138$, Doorhandle: adj. $R^2 = 0.088$) for the two activities with the highest R^2 . A small random jitter was added to the window sizes for better visibility.

Activity name	Description	Sequence	Activity type	Accuracy	AUC	Num pos	Pos %
Dwarf	Walking in squat position	From the first step till the last step; if child stops coding is paused	Playful	0.9512	0.8838	1113	5.09
Rabbit	Legs between hands, first moving forward hands, then jumping with legs	Begins with hand stretching, ends with completing the jump	Playful	0.9790	0.8711	362	1.66
Book	Turning pages in a book	From grasping a page till releasing it	Everyday	0.9571	0.8667	750	3.43
Nose	Touch nose	From reaching towards nose till releasing it	Everyday	0.9797	0.8576	503	2.30
Light on	Turning on the light, then releasing hands	From lifting hand till the hand is next to the body in the start position	Everyday	0.9800	0.8558	375	1.72
Door handle	Grabbing door handle, pushing down, releasing hands	From lifting hand till releasing the door handle	Everyday	0.9686	0.8451	645	2.95
Peck	Peck the skin on the back of the hand	From reaching towards the hand till releasing it	Everyday	0.9710	0.7892	535	2.45
Frog	Jumping with open legs, hands in the air beside the body	From moving upward till arriving to the lowest point	Playful	0.9780	0.7781	346	1.58
Glass grabbing	Grabbing a glass	From reaching glass till grabbing	Everyday	0.9828	0.7696	268	1.23
Pudding eat	Eating a pudding	One mouthful, from putting spoon into the mouth till moving it away from mouth	Everyday	0.9598	0.7678	836	3.82
Clapping	Clapping hands	From approximating hands till start position	Everyday	0.9825	0.7577	208	0.95
Drinking	Drinking from a glass	From touching mouth with the glass till releasing it	Everyday	0.9630	0.7473	178	0.81
Glass lifting	Lifting a glass	From grabbing till lifting the upmost point	Everyday	0.9802	0.6961	216	0.99
Sock on (same side)	Put on socks	From grasping a sock till the sock is on the foot	Everyday	0.9752	0.6889	512	2.34
Sock on (other side)	Put on socks	From grasping a sock till the sock is on the foot	Everyday	0.9697	0.6863	462	2.11

Table 2. The Budapest activity test battery (BATB) and performance of the machine learning methods. The table was sorted by AUC with the highest recognition figures at the top. Num pos is the number of occurrences of a given activity in the dataset.

The categorization performance is better at higher AUC figures, ranging from 0.5 as minimum value to 1.0 as the maximum. However, the closer AUC is to the maximum value the lower is the variance of the AUC and as a consequence the window size has lower impact on the performance.

Our results enable the development of game applications which improve motor skills as the ML model integrated into the Senskid system is able to give feedback to the user about the accuracy of movement execution. At least the 17 movement types can be used in such applications as these were identified with very high AUC values. We achieved the highest AUC values mainly for the playful activities which children frequently execute during preschool or school activities or physical training (e.g. crab and spider crawling, bear walking, etc.). Therefore, our activity battery can be easily applied in preschools and schools with the help of the teachers.

Activity name	Description	Sequence	Activity type	Accuracy	AUC	Num pos	Pos %
Toothbrush (other)	Put toothpaste on to the toothbrush	From grabbing toothbrush and toothpaste till finishing putting toothpaste on the toothbrush	Everyday	0.9720	0.6855	371	1.69
Hand wash	Washing hands	From rubbing hands till hands are under water	Everyday	0.9814	0.6849	285	1.30
Snack eat	Eating a snack	One bite, from putting snack into the mouth till moving away from mouth	Everyday	0.9677	0.6587	288	1.32
Knee (same)	Touch knee	From reaching towards knee till releasing it	Everyday	0.9762	0.6478	227	1.04
Knee (other)	Touch knee	From reaching towards knee till releasing it	Everyday	0.9171	0.6161	134	0.61
Shoe on (same side)	Put on shoes	From grasping a shoe till the shoe is on the foot	Everyday	0.9734	0.6139	417	1.91
Pray	Hands together in pray style	From approximating hands till releasing them	Everyday	0.9740	0.5993	389	1.78
Toothbrush (same)	Put toothpaste on to the toothbrush (same: pushing toothpaste with the hand smartwatch on it)	From grabbing toothbrush and toothpaste till finishing putting toothpaste on the toothbrush	Everyday	0.9534	0.5637	226	1.03
Glass to mouth	Lifting glass to the mouth	From lifting glass to the mouth till it touches the mouth	Everyday	0.7987	0.5631	31	0.14
Shoe off (other side)	Take off shoes, opposite side of the body relative to the hand with the smartwatch	From grasping a shoe till the shoe is off the foot	Everyday	0.8706	0.5383	53	0.24
Shoe on (other side)	Put on shoes	From grasping a shoe till the shoe is on the foot	Everyday	0.9387	0.5366	247	1.13
Sock off (same side)	Take off socks	From grasping a sock till the sock is off the foot	Everyday	0.9280	0.5110	167	0.76
Sock off (other side)	Take off socks	From grasping a sock till the sock is off the foot	Everyday	0.8481	0.5	115	0.53
Shoe off (same side)	Take off shoes, same side of the body relative to the hand with the smartwatch	From grasping a shoe till the shoe is off the foot	Everyday	0.8130	0.5	70	0.32

Table 3. The Budapest activity test battery (BATB) and performance of the machine learning methods. The table was sorted by AUC with the highest recognition figures at the top. Num pos is the number of occurrences of a given activity in the dataset.

Game applications relying on ML algorithms which request and measure these activities could be used in such context to facilitate and improve movement, or predict motor or neurocognitive problems. Although some of the everyday activities were less successfully recognized in the present study, this could be improved by increasing sample size of the collected data. Many children with motor/neurocognitive problems³⁴ have difficulties with performing everyday activities, thus, there is a huge need for gamified technological solutions which improve their motor skills.

Methods

Participants. Thirty-nine children participated in the study (22 girls, 17 boys; age range: 6.59–8.38; median age = 7.45), but the data of 5 children was not used in the analysis due to technical problems (either the movement data or the video recording was lost). Thus, finally we analyzed the data of 34 children (19 girls; 15 boys age range: 6.59–8.38; median age = 7.47). All children were typically developing first graders from three elementary schools (see Sect. 1.3). The parents of all participants gave their written informed consent to the study.

Materials. *Budapest activity test battery (BATB).* The Budapest Activity Test Battery (BATB) was developed by our research group, and it contains complex activities (the activities and their definitions are presented in Tables 1, 2, 3, that are executable by 6–8-year-old children. The selection criteria for the activities were that they should be motivating enough for the children to be integrated later into a game application, they should improve different motor skills after regular practice (fine and gross motor activities, arm-leg coordination, cross-side and same-side activities), and they could be either shown by a character on the mobile/tablet's screen or explained (asked) verbally. Therefore, two main types of activities were included: playful, entertaining (mostly reproducing or mimicking movements of animals) activities that the child would be motivated to perform, and everyday activities that (being part of a child's everyday routine) the parent would be motivated to ask from the child (e.g. taking off the shoes, washing hands).

We used different instructional methods for the playful and everyday activities. Playful activities (N = 10) were shown by the experimenter, and children were asked to imitate them. In contrast, children were instructed verbally to perform everyday activities (N = 24). If they were not able to carry them out, then the experimenter demonstrated the actual activity. The children were asked to repeat each of them 5 times in order to obtain more data. Some of the activities required some equipment e.g. ball, book, building blocks, glass, spoon, snack, toothbrush, toothpaste.

Playful activities (except for hopscotch) were repetitive movement sequences: the same movement elements were repeated over a predetermined distance (e.g. crawling like a crab across the room). Everyday activities were single (not repetitive) action sequences (their units were functional on their own, e.g. throwing up a ball,

grabbing the door handle). At the end of the data collection the number of the activities varied from 31 to 2656 (number of positive, ground truth).

Device and software. The data collection equipment consisted of two devices, one sensor device (smartwatch) and another one (smartphone) for controlling the sensor device and managing sensor data, both running SensKid software. We used Apple Watch as sensor device which is a commercially available Apple product, regarding product details, please see <https://www.apple.com/>.

SensKid software is a member of the SensX software family, which is under development and not yet commercially available. The sensor device contained a 9-axis motion sensor (3-axis gyroscope + 3-axis accelerometer + 3-axis magnetometer) and samples data at 50 Hz (50 sample/sec). Each sample data point contained 3 dimensional parameters of the device except attitude which had 4 dimensions (9 + 4). During the experimental session the sensor device processed and stored the gyroscope and accelerometer data in real time. At the end of the session the processed sensor data was sent to the measurement device connected via Bluetooth that in parallel recorded the session on video, then transferred the data and the video to our network servers. The synchronisation of the raw data and the video was made automatically by the SensKid software.

Procedure. Data collection took place in classrooms/gyms of three elementary schools (Kispesti Vass Lajos Általános Iskola: N = 8; Virányos Általános Iskola: N = 12; Terézvárosi Két Tannyelvű Általános Iskola: N = 19). The school psychologist/ class teacher was contacted first directly via email, then, an informed consent was asked from the director of the school. The class teacher helped us in contacting the parents to obtain their informed consents. All methods were performed in accordance with the relevant guidelines and regulations, and all the experiment protocol for involving humans was in accordance to 2018 Declaration of Helsinki.

Children were tested in groups of 2 or 3 in the presence of their teacher. The experimenter informed the child about what would happen in the test and put the sensor device on the wrist of the child, on the child's dominant hand (which was determined by offering a high five to the child and checking which hand he or she used spontaneously). The experimenter then set the connection between the smartwatch and the smartphone and launched the recording.

The order of the activities was not fixed, it was determined ad hoc, based on (1) what the child wanted to perform, (2) how exhaustive the activity was (e.g. after 4–5 playful activities, children got tired so we changed to a less exhaustive everyday activity), (3) how much space was available for the activities. Additionally, most children did not perform all activities because of time limitation, exhaustion or other personal reasons (e.g. the child did not want to or was not allowed to perform something, e.g. eating the snack). Children were asked to repeat an activity two or five times (two times in case of longer sequences, like bear walk or frog jump, and five times in case of shorter activities like switching the light on/off). The total number of the collected samples per activity is reported in Tables 1, 2, 3.

The study was approved by the Unified Psychological Research Ethical Committee (EPKEB; reference number: 2019/18).

Video coding. For video coding we used Solomon Coder (© András Péter). The coding protocol included definitions of the 40 activities and the length of an activity sequence (bout length: start—endpoint, without interruption). Per definition we used the number of positives as the number of occurrences of a given activity, which means the number of bouts, activity running ongoing without being interrupted by any other activities. For the activities and their definitions, see Tables 1, 2, 3. Video recordings were coded by five coders. Video recordings were manually synchronized to the inertial data. All coders were trained using a standardised protocol of the department, and inter-coder reliability analyses were performed during training to ensure consistent labelling.

Data analysis. We chose LGBM for the categorisation (LGBM 3.1.1, <https://pypi.org/project/lightgbm/>), because our previous research (submitted) on other datasets showed that LGBM did just deliver the best performance, but it significantly over-performed other boosting methods (e.g. XGBM) in speed and computational efforts (1.2 h vs 8.5 h per iteration).

The hyperparameters were set to default, except for that we used the 'unbalance = true' parameter, as our dataset is unbalanced as it's expected. To evaluate the machine learning model, we separated the data to independent data sets for training, validating and testing. This has been carried out by k-fold cross-validation (CV). In k-fold CV, the training data is randomly partitioned in k equal subsets. The model is then trained on k – 1 subsets, and the remaining one is used for validation. We used a threefold cross validation, splitting the dataset 25 + 25% for train and 25% for validation group. We used the validation group to tune the hyper-parameters and check the overfitting. Then all the remaining 25% of the dataset made up the test group. We did the cross-validation per child, so the training and test sets did not contain the data of the same subject. Therefore, this method is referred to as subject-independent CV.

One of the most important statistical assumptions for ML processes is that samples are independent and identically distributed, that is, all the data points are sampled independently from the same distribution. However, samples drawn from the same subject are most likely not independent. This means that the similarity of samples drawn from the same participants is likely to be higher than that of samples drawn from different participants (see also³⁵).

This kind of bias of k-fold CV may overestimate the performance of categorization. This problem of k-fold CV is more serious when it is used with overlapping sliding windows, as in our experiment, because these overlaps between adjacent windows are another source of unwanted dependency between data points. To address these issues, the training and testing sets should be split by participants. According to this method, which is known

as subject-independent CV, in each iteration the model is trained on all the participants except those, which are used for testing. In our case, we separated the participants into 3-folds using 2-folds for training and one-fold for testing.

We used the dynamic overlapping sliding window technique for segmentation of the data with 5 sampling unit shifts. Sliding window sizes of 15, 32, 60, 81, 100, 149 sample points (0.3, 0.6, 1.2, 1.6, 2.0, 2.9 s) was considered for feature computation; this provided sufficient temporal resolution of activity and was short enough to capture bouts of activities with the shortest duration. Successive windows had an overlap of 5 sample points. Windows containing transitions between different activities were labelled as the activity at the end of the bout. Thus each window contained activity data corresponding to exactly one video-labelled activity.

As any multi-class problem could be built up from binary classifications, we decided to use separate binary models, not one multi-label, one for each activity, and comparing the positive class to the remaining 39 activities. We ran 20 iterations per window size per fold, 60 total runs for every activity and calculated the weighted AUC value (0.5–1.0) of the run of the activity as the indicator of recognition success. We used the feature set as published earlier in³⁵.

Conclusion

In summary, we collected activity motion data with a special SensKid software by means of wearable smart-watches on the children's wrist, asking them to show various kinds of daily routine or playful activities. We analysed the data of 34 children who were typically developing first graders from three elementary schools. Our aim was to build a machine learning model which could recognize these activities.

Light Gradient Boosted Machine (LGBM) learning algorithm was used, with a threefold cross validation in a binary classification task. We used sliding window technique during the signal processing, and we also analysed the effect of window size for the analysis of each behaviour element to achieve the most effective settings. Seventeen activities out of 40 were successfully recognized with AUC values above 0.8.

In summary, the LGBM is a very promising solution for recognizing daily routine or playful activities among children in real life situations, which is not sensitive to the window size. Big advantage of our finding that this machine learning method even works on commercially available devices, which could open the window for more promising examination of children behaviour in every-day situations.

Received: 13 November 2021; Accepted: 17 March 2022

Published online: 31 March 2022

References

- Bateson, M. & Martin, P. *Measuring Behaviour: An Introductory Guide* (Cambridge University Press, 2021).
- Elliott, K. H., Le Vaillant, M., Kato, A., Speakman, J. R. & Ropert-Coudert, Y. Accelerometry predicts daily energy expenditure in a bird with high activity levels. *Biol. Lett.* <https://doi.org/10.1098/rsbl.2012.0919> (2013).
- Wang, Y. *et al.* Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. *Mov. Ecol.* <https://doi.org/10.1186/s40462-015-0030-0> (2015).
- Airaksinen, M. *et al.* Automatic posture and movement tracking of infants with wearable movement sensors. *Sci. Rep.* **10**, 1–13. <https://doi.org/10.1038/s41598-019-56862-5> (2020).
- Gao, L., Zhang, G., Yu, B., Qiao, Z. & Wang, J. Wearable human motion posture capture and medical health monitoring based on wireless sensor networks. *Meas. J. Int. Meas. Confed.* **166**, 2. <https://doi.org/10.1016/j.measurement.2020.108252> (2020).
- Gerencsér, L., Vásárhelyi, G., Nagy, M., Vicsek, T. & Miklósi, A. Identification of behaviour in freely moving dogs (*Canis familiaris*) using inertial sensors. *PLoS ONE* **8**, e77814. <https://doi.org/10.1371/journal.pone.0077814> (2013).
- Hadjidj, A., Souil, M., Bouabdallah, A., Challal, Y. & Owen, H. Wireless sensor networks for rehabilitation applications: Challenges and opportunities. *J. Netw. Comput. Appl.* **36**, 1–15. <https://doi.org/10.1016/j.jnca.2012.10.002> (2013).
- Porciuncula, F. *et al.* Wearable movement sensors for rehabilitation: A focused review of technological and clinical advances. *PM R* **10**, S220–S232. <https://doi.org/10.1016/j.pmrj.2018.06.013> (2018).
- Grimm, B. & Bolink, S. Evaluating physical function and activity in the elderly patient using wearable motion sensors. *EFORT Open Rev.* **1**, 112–120. <https://doi.org/10.1302/2058-5241.1.160022> (2016).
- Najafi, B., Armstrong, D. G. & Mohler, J. Novel wearable technology for assessing spontaneous daily physical activity and risk of falling in older adults with diabetes. *J. Diabetes Sci. Technol.* **7**, 1147–1160. <https://doi.org/10.1177/193229681300700507> (2013).
- Jalloul, N. Wearable sensors for the monitoring of movement disorders. *Biomed. J.* **41**, 249–253. <https://doi.org/10.1016/bj.2018.06.003> (2018).
- Konok, V., Bunford, N. & Miklósi, Á. Associations between child mobile use and digital parenting style in Hungarian families. *J. Child. Media* **14**, 91–109. <https://doi.org/10.1080/17482798.2019.1684332> (2020).
- Hotu, S., Carter, B., Watson, P. D., Cutfield, W. S. & Cundy, T. Increasing prevalence of type 2 diabetes in adolescents. *J. Paediatr. Child Heal.* **40**, 201–204. <https://doi.org/10.1111/j.1440-1754.2004.00337.x> (2004).
- Skinner, A. C., Perrin, E. M. & Skelton, J. A. Prevalence of obesity and severe obesity in US children, 1999–2014. *Obesity* **24**, 1116–1123. <https://doi.org/10.1002/oby.21497> (2016).
- Marsden, N., Wollmann, T., Lohmann, B. & Meixner, G. Formative evaluation of smartwatch exergaming. *Mensch und Comput. 2015 - Work*. 145–146, DOI: <https://doi.org/10.1515/9783110443905-021> (2015).
- Lyons, E. J. Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards. *Games Heal. J.* **4**, 12–18. <https://doi.org/10.1089/g4h.2014.0072> (2015).
- Gall, M. *et al.* A novel approach to assess sleep-related rhythmic movement disorder in children using automatic 3D analysis. *Front. Psychiatry* **10**, 1–10. <https://doi.org/10.3389/fpsy.2019.00709> (2019).
- Kashi, S., Feingold-Polak, R., Lerner, B., Rokach, L. & Levy-Tzedek, S. A machine-learning model for automatic detection of movement compensations in stroke patients. *IEEE Trans. Emerg. Top. Comput.* <https://doi.org/10.1109/TETC.2020.2988945> (2020).
- Lorenzi, P., Rao, R., Romano, G., Kita, A. & Irrera, F. Mobile devices for the real-time detection of specific human motion disorders. *IEEE Sensors J.* **16**, 8220–8227. <https://doi.org/10.1109/JSEN.2016.2530944> (2016).
- Jaiswal, S., Valstar, M. F., Gillott, A. & Daley, D. Automatic Detection of ADHD and ASD from Expressive Behaviour in RGBD Data. Proc. - 12th IEEE Int. Conf. on Autom. Face Gesture Recognition, FG 2017 - 1st Int. Work. on Adapt. Shot Learn. for Gesture Underst. Prod. ASL4GUP 2017, Biom. Wild, Bwild 2017, Heterog. Face Recognition, HFR 2017, Jt. Chall. on Dominant

- Complementary Emot. Recognit. Using Micro Emot. Featur. Head-Pose Estim. DCER HPE 2017 3rd Facial Expr. Recognit. Analysis Challenge, FERA 2017 762–769, DOI: <https://doi.org/10.1109/FG.2017.95> (2017). 1612.02374.
21. Zwaigenbaum, L. *et al.* Early Intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics* **136**, S60–S81. <https://doi.org/10.1542/peds.2014-3667E> (2015).
 22. Ardalan, A., Assadi, A. H., Surgent, O. J. & Travers, B. G. Whole-body movement during videogame play distinguishes youth with autism from youth with typical development. *Sci. Rep.* **9**, 1–11. <https://doi.org/10.1038/s41598-019-56362-6> (2019).
 23. Min, C. H. Automatic detection and labeling of self-stimulatory behavioral patterns in children with Autism Spectrum Disorder. *Proc. Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc. EMBS* 279–282, DOI: <https://doi.org/10.1109/EMBC.2017.8036816> (2017).
 24. Rad, N. M. *et al.* Convolutional neural network for stereotypical motor movement detection in autism. *Biol. Lett.* **1511**, 01865 (2015).
 25. Ofcom. Children and parents: media use and attitudes report Content consumption and online activities. *OFCOM* (2021).
 26. Schneider, H. & Eisenberg, D. Who receives a diagnosis of attention-deficit/hyperactivity disorder in the United States elementary school population?. *Pediatrics* <https://doi.org/10.1542/peds.2005-1308> (2006).
 27. Huang, W. *et al.* Shallow convolutional neural networks for human activity recognition using wearable sensors. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2021).
 28. Dehghani, A., Sarbishi, O., Glatard, T. & Shihab, E. A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors. *Sensors* **19**, 10–12. <https://doi.org/10.3390/s19225026> (2019).
 29. Bersch, S. D., Azzi, D., Khusainov, R., Achumba, I. E. & Ries, J. Sensor data acquisition and processing parameters for human activity classification. *Sensors (Switzerland)* **14**, 4239–4270. <https://doi.org/10.3390/s140304239> (2014).
 30. Banos, O., Galvez, J. M., Damas, M., Pomares, H. & Rojas, I. Window size impact in human activity recognition. *Sensors (Switzerland)* **14**, 6474–6499. <https://doi.org/10.3390/s140406474> (2014).
 31. Ni, Q. *et al.* Leveraging wearable sensors for human daily activity recognition with stacked denoising autoencoders. *Sensors (Switzerland)* **20**, 1–22. <https://doi.org/10.3390/s20185114> (2020).
 32. Chung, S., Lim, J., Noh, K. J., Kim, G. & Jeong, H. Sensor data acquisition and multimodal sensor fusion for human activity recognition using. *Sensors (Switzerland)* <https://doi.org/10.3390/s19071716> (2019).
 33. Awais, M. *et al.* Performance evaluation of state of the art systems for physical activity classification of older subjects using inertial sensors in a real life scenario: A benchmark study. *Sensors* <https://doi.org/10.3390/s16122105> (2016).
 34. Ahmadi, M. *et al.* Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. *J. NeuroEng. Rehabil.* **15**, 105. <https://doi.org/10.1186/s12984-018-0456-x> (2018).
 35. Ferdinandy, B. *et al.* Challenges of machine learning model validation using correlated behaviour data: Evaluation of cross-validation strategies and accuracy measures. *PLoS ONE* **15**, 1–14. <https://doi.org/10.1371/journal.pone.0236092> (2020).

Acknowledgements

Our research was supported also by the Innovation Office (ELTE Thematic Excellence Programme 2020, TKP2020-IKA-05; OTKA K124458; OTKA KH129603; OTKA K135478; OTKA PD134984), by the Ministry for Innovation and Technology (ÚNKP-21-5 New National Excellence Program) and the Hungarian Academy of Sciences (MTA 01 031; Bolyai Janos Research Fellowship; MTA Lendület Programme, #LP 2018-3/2018).

Author contributions

G.C. made the machine learning models and analysed the results and wrote the main manuscript text, K.L.P. and V.K. conducted the experiment(s) and contributed in writing the manuscript, B.F. prepared the linear regression analysis and the corresponding figures. M.A. contributed in writing the manuscript. All authors reviewed the manuscript.

Funding

Open access funding provided by Eötvös Loránd University.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022