Non-intrusive and Privacy Preserving Activity Recognition System for Infants Exploiting Smart Toys

Niko Bonomi and Michela Papandrea(✉)

University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Lugano, Switzerland
{niko.bonomi,michela.papandrea}@supsi.ch

Abstract. The Human Activity Recognition (HAR) research area showed great advances in the last decade, achieving excellent prediction performances and great applicability, which is reflected on the wearable sensors market adoption. However, most of the research effort concentrated on an adult target population. When considering a younger population of infants or children, currently available HAR solution based on wearable devices are not applicable anymore. In this paper we present an HAR based solution targeting infants, based on a non-intrusive and privacy-preserving measurement methodology which allows the preservation of children behaviour and the collection of objective data (particularly important for clinical observation purposes). The proposed solution, based on the usage of a set of smart toys (AutoPlay toys-set) achieves great performances in the recognition of a set of 12 toy-activity pairs, reaching accuracy values up to 96%. These results pave the way to a broad application of the presented methodology on objective analysis of humans motor skills.

Keywords: Activity · Motricity · Infants · Neurodevelopment · Toys · HAR · Infants activity recognition

1 Introduction

In the last decade, thanks to the miniaturization of hardware components, it has been quite easy to equip almost every digital device with some sort of sensors and actuators that coupled with transmission capabilities enable an endless number of possibilities. One of such possibility is Human activity recognition (HAR), which nowadays is widespread in everyday objects, like smartwatches or fit-tracker. HAR main goal is to identify and interpret the actions executed by a human based on some measured sensor data. HAR can be performed using information coming from different sources like smartphone or wearable sensors, camera footage, LIDAR technology: those observations are feed to an intelligent algorithm that performs a prediction based on supervised knowledge.

S. Spinsante et al. (Eds.): HealthyIoT 2021, LNICST 432, pp. 3–18, 2022.
https://doi.org/10.1007/978-3-030-99197-5_1

The HAR research area showed great advances in the last decade, achieving excellent prediction performances and great applicability, which is reflected by the growth of the wearable sensors marked. However, most of the research effort concentrated on the adult target population. When considering a younger population of infants or children, currently available HAR solutions based on wearable devices are not applicable anymore [7]. Wearable sensors introduce a bias in the young Human Activity Recognition (yHAR) systems: children tend to focus their attention on the sensor device itself hence their behavior is not natural. Solutions based on cameras recordings are adopted instead, reaching not very accurate results on very small children (i.e., infant skeleton extraction from videos is still a challenge) and dealing with the issue of privacy.

AutoPlay [4] is an innovative project which deals with the challenges of young Human Activity Recognition: it adopts the aforementioned HAR technologies in the field of infants neuro-development monitoring. The AutoPlay general goal is to anticipate the diagnosis of autism spectrum disorders (ASD), neuro developmental disorders and other social fragilities. The AutoPlay methodology is based on a toys kit, a set of smart toys equipped with sensors, for infants motor skills interpretation. The sensors allow the collection of inertial data (e.g., accelerometer, gyroscope and magnetometer). The usage of such smart-toys provides an effective measurement methodology when applied to very small children: this is to handle the important issue present when dealing with the measurement of motricity of children which are very small, or very sensitive at a sensory perception level. It does not provide an obstacle to the freedom of movement when compared to bulky wearable sensors. And, additionally, it does not affect the privacy of small infants, as opposed to video recording and computer vision based methodologies. In this paper we present the results of applying current HAR methodologies exploiting the AutoPlay smart toys, as non-worn sensors.

The results obtained and presented in this work are encouraging and demonstrating that nowadays available technology is ready for such application, paving the way for a broad adoption of non intrusive and privacy preserving methodologies for the observation and monitoring of infants, and more in general for fine motor-skills measurement.

In this paper we provide a brief description of the related SoA (Sect. 2). Successively, Sect. 3 describes the methodologies we exploited to collect training data for the predictive models. In Sect. 4 we describe the learning approach and, finally, in Sect. 5 we describe the obtained results, distinguishing between in-lab and real use-case scenarios.

2 Related Work

Nowadays, human activity recognition (HAR) services are embedded in almost every wearable smart devices (e.g., smart watches, fitness trackers, etc.) and companion devices such as smartphones. The main objective of such embedded HAR services is the identification of humans movements and actions, and the recognition of behavioral patterns on the base of heterogeneous measured data.

There exist various types of wearable sensors, meant for being worn on different parts of the human body. Those kinds of sensors might be uncomfortable to be worn, this means that they are not a good solution for long term monitoring of human activities. In the last decade the smartphone industry evolved, allowing the embedding of those kinds of sensors directly on a portable device that almost everybody carries in their pocket. This kind of sensors bring up new research opportunities for human-centered activity recognition. They usually embed three axis accelerometer, gyroscope, microphone, camera and many more sensors depending on the vendor. Nowadays almost every smartphone is equipped with such sensors that coupled with smart algorithms provide the user useful information such as the daily amount of steps, workout session duration and also some more interesting metrics like the sleep quality.

From the related literature, Anguita et al. [1] collected and shared a dataset of human activities from various persons and annotated in each time unit the activity that the person was carrying out: this allowed the generation of a ground truth where the data is connected to a human activity. The generated dataset contains accelerometer and gyroscopic sensors data collected from smartphones. Exploiting this dataset they have trained a support vector machine classifier using a rolling window of about 2 s and a step of 50% of the size of the window, and achieved an activity prediction accuracy score of 74%. Also Papandrea et al. [6] have applied such methodology in a location prediction and mobility modelling system, with the main goal of cutting computational costs and increasing the prediction performances based on a personalization strategy. Also in this case they have applied an averaging moving window to compute the necessary features, and reached an accuracy score of 94% in the prediction of 9 different activities.

The research study presented in this paper concerns the HAR methodologies applied to toddlers, and more in particular with the support of augmented toys. Numerous works in literature have already exploited this topic, Rivera et al. [7] presented an architecture of augmented toys, consisting of smart cubes that are able, using a set of light sensors, to detect how they are oriented and how they are placed with respect to other cubes. They share the observation related to the issue of children wearing conventional wearable sensors, stating that this could alter children behaviour and cause distractions, thus invalidating the measurements. This is actually the main reasons behind the practice of embedding sensors inside common objects, like toys, so that the children are not disturbed and biased by them.

An interesting work carried out by T. L. Westeyn et al. [11] in 2010, presents a toys-kit that helps in the annotation process of children activities trough cameras and sensors embedded within toys. The author proposed a specific tool that exploits smart toys as an assistant to a video-monitoring system for children and adults. They presented good results obtained on the gathered data, performing a support vector machine based classification to classify the kind of activity played by the child (i.e., jumps, shake, spin etc.).

Many different Inertial Measurement Units sensors are available on the market at affordable prices, and with an acceptable measurement resolution for HAR application. Among these, the Shimmer IMU is a sensor device widely exploited among different research projects, especially medical applications. Mehmood et al. [5] exploited this sensor in the field of human activity recognition. They have obtained some encouraging results with sampling frequency 50 Hz, demonstrating it allows the collection of all relevant information about human activities. Using a wide set of classification algorithms they have discovered that the *support vector machine* performs very well for the recognition of stationary activities like sit, lying down, stand still, etc., meanwhile the *random forest* algorithm reaches great results on non-static activities recognition like running, walking, jumping etc. In the field of activity recognition, an interesting work has been carried out by [2] Antar et al.; it showcases the challenges in the field of HAR based on wearable sensors, providing a complete description on how to perform an initial Exploratory Data Analysis (EDA), on which filters needs to be used during the preprocessing phase, on how to perform an effective segmentation of the data and on which are the features that are worth considering in the case that the data will be feed into an intelligent algorithm.

With this paper we intend to advance the current SoA presenting the application of supervised technologies to a more challenging research problem, the *indirect infant activity recognition*. With the proposed approach we aim to provide new tools to the infants and childhood health research area, more specifically with the goal of anticipating ASD diagnosis.

3 The Dataset

A set of measurement sessions have been carried out to sample inertial data: each measurement consists in a play session exploiting the AutoPlay toys kit [4] and collects a 9-dimentional raw dataset consisting of *3D acceleration low noise*, *3D acceleration wide range* and *3D gyroscope data* . Each toy is tailored to accommodate one or more IMU sensors.

The sensor node of choice is the Shimmer IMU unit which is equipped with multiple sensors: low-noise and wide range accelerometer, gyroscope, magnetometer, humidity and temperature sensor and altimeter, allowing to sample data at a frequency up to 2048 samples per second [9].

The measurement sessions have been enriched with the help of cameras that enable the annotation of the activities carried out by the human, thus allowing to create a supervised dataset for the training of activity prediction algorithms. The cameras recorded videos at 25 fps: the synchronization between the collected data and the recorded videos have been performed with a two-phases synchronization methodology, presented in [8]. Each session was manually post processed in order to associate a ground truth to the data: each frame of a video is associated with an activity, and each sample of the collected data is associated to a frame (hence to an activity as well).

3.1 Synthetic Dataset: In-Lab Data Collection

A first dataset was collected directly in lab, allowing for an easy data acquisition process necessary to create a labeled training dataset.

The dataset includes samples related to three toys: an elephant, a small ball and a car (Fig. 1). The elephant and the ball had one sensor node mounted inside the toy, meanwhile the car has two sensors mounted directly into two wheels (one in the front, one in the back of the car) allowing the independent measurement of both front and rear wheels movements.

Fig. 1. Toys used in lab

The dataset contains a set of (different) activities per toy. In total we collected samples belonging to 12 different couples <toy, activity> (list of activities per toy is reported in Table 1). The listed activities have been identified to be the most representative ones for toddler (9–24 months) in terms of frequency of appearance, as observed during a real world data gathering sessions involving children. Our goal was to select a reduced set of significant activities, which are interesting from an activity recognition point of view, and feasibly implementable in a more realistic scenario.

Table 1. Toys activities

Car	Ball	Elephant
Drive	Toss	Let it fall
Overturn	Roll over	Overturn
Turn wheel	Shake	Throw
Knock	Turn	Knock

The in-lab data collection involved two adult persons: each of them performed a data collection session of 7 min for each pair <toy, activity>. This resulted in a total of 168 min of data collected in lab for the synthetic settings.

An additional synthetic data measurement session was performed to sample a mixed sequence of activities, where each involved person simulated the movements of a toddler playing with the exploited toys. Each person had the possibility to freely choose the amount of time to dedicate to each activity (in the list shown above) and the order of the activities in the sequence. This resulted in a total amount of 12 min of collected data.

3.2 The AutoPlay Dataset

In the scope of the AutoPlay project, a real world dataset has been gathered. Data from a vast variety of toddler have been sampled using the AutoPlay augmented toys. Each measurement session has also been recorded with cameras for annotation purposes. The collected sensor data has been synchronized with the captured video and thus with the annotated ground truth thanks to the methodology presented in [8]. This work proposed an approach for two main problems: the data synchronization issue, due to the lack of synchronization between the cameras used to record the toddler playing and the actual sensors installed into the toys (without an on-board real time clock available), and the data-time alignment due to the sensor clock drift problems.

3.3 Annotation

The annotation procedure have been possible thanks to the footage captured by various cameras installed on the measurement environment. Each video is then post-processed manually and each frame of the video is associated with an activity declared in a specific pool of possible activities selected a priori. In the case of the in-lab data acquisition this pool of possible activities was reduced to 4 per each toy (as mentioned above) meanwhile the AutoPlay dataset is comprehensive of a large set of defined activities which can be categorized in macro groups.

- **Functional**: A functional activity is described as an activity that can enable another subsequent set of activities. The functional activity is used to achieve an objective such as stack some toys to build a tower or push away a non desired toy.
- **Exploratory**: An exploratory activity allows, as suggested by the term, the toddler to explore both the environment and the toys properties stimulating the 5 senses.
- **Rotation**: This is a category of activities in which the toddler performs some kind of rotation of the toy with an exploitative purpose, of as a functional activity.

Table 2 shows the complete list of possible activities identified for the AutoPlay dataset.

Table 2. Complete list of infants play activity of reference

Activity	Category
Push	Functional
Shift	Functional
Lay	Functional
Lift	Functional
Lower	Functional
Drag	Functional
Throw	Functional
Tender	Functional
Grab	Functional
Knock over	Functional
Stack	Functional
Pick	Exploratory
Hold in hand	Exploratory
Let it fall	Exploratory
Shake	Exploratory
Hit	Exploratory
Bite	Exploratory
Knock	Exploratory
Turn	Rotation
Overturn	Rotation
Roll over	Rotation

4 Predictive System

In order to build a predictive methodology able to identify, given raw inertial data (as described in Sect. 3), the related toddler activity we realized a data workflow which includes the following steps:

1. data collection (described in Sect. 3)
2. data annotation (described in Sect. 3.3);
3. data synchronization (as described in [8]);
4. features extraction;
5. activity prediction model training and parameter tuning;
6. model validation on real use-case data.

Regarding the *features extraction* phase (point 4 of the workflow), we calculated a feature vector of 14 variables per each dimension of the input raw data, including standard statistical features like: mean, standard deviation, max, min and signal vector amplitude (as presented in [1]).

Table 3. Feature vector of 14 variables calculated per each raw input data dimension

Feature	Description
min	Min value
max	Max value
mean	Signal mean
mad	Median absolute value
std	Standard deviation
pow	Signal power
skew	Skewness
kurtosis	Kurtosis
deriv mean	mean of the derivative
deriv std	Standard deviation of the derivative
SMA	Signal magnitude area
entropy	Signal entropy
iqr	Interquartile range
snr	Signal to noise ratio

The 9-dimentional raw data (acceleration low noise 3DoF, acceleration wide range 3DoF, giroscope 3DoF) is enriched with the *magnitude* of the acceleration vector, generating a 10-dimensional features vector. Among the 14 variables shown in Table 3, only 13 of them are calculated over all the 10 raw dimensions. The SMA variable is computed using Eq. 1 over the 3-dimensional axis of the acceleration low noise. Hence the final resulting features vectors belong to a 131-dimensional space.

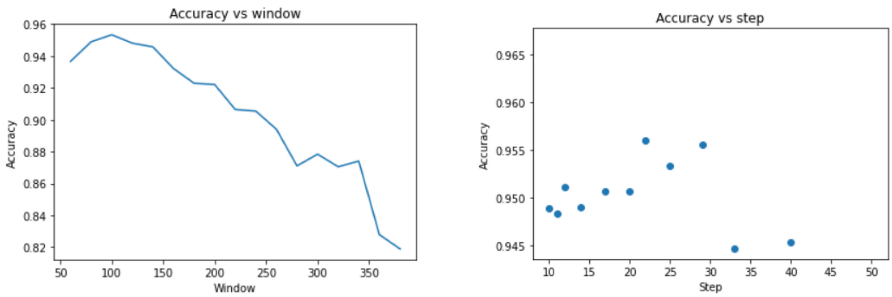
$$SMA = \frac{\sum_{n=1}^{N_x} |x_n| + \sum_{n=1}^{N_y} |y_n| + \sum_{n=1}^{N_z} |z_n|}{N_{samples}} \tag{1}$$

The features described above have been calculated per each raw sample data. As part of the features extraction phase, we calculated the features over a temporal window. In order to generate the training set for our activity prediction model we extracted aggregated features over the above mentioned temporal window, sliding it over time; in particular the window is sliding over the temporal signal of a fixed sized step. In the following sections we describe the study performed in order to identify the optimal values for both *window size* and *sliding step length*. The objective of this study is to find the smallest possible windows size and step length necessary to carry detailed information about the activity performed, without compromising the performances of the activity recognition classifier. The objective related to the size of the window, is driven by the necessity of being responsive in the recognition of the activity, in a context where the mean duration of an activity is the order of seconds.

In the first instance we performed a grid search for the window size optimal value, fixing the sliding step length to 25% of the window. The metric used to evaluate the performances of the activity classification model per each window size value, is the *accuracy*. This procedure has been performed separately per each toy, because the average infant activity duration strictly depend on the exploited toy.

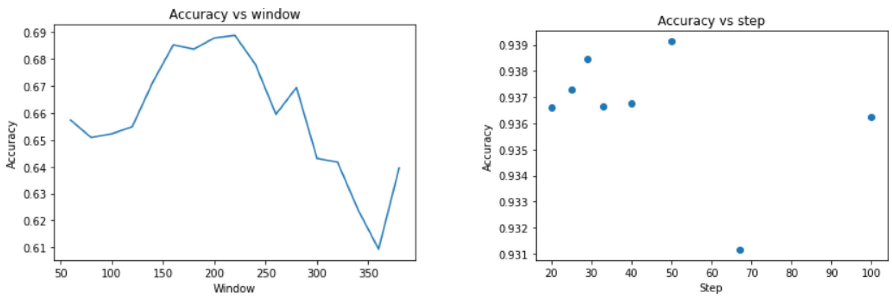
For what concern the **ball** toy, as shown in Fig. 2a, the best window size obtained is identified by the global maximum of the window size versus accuracy graph, in the searched range: the optimal value found correspond to 100 samples, that means 1 s window size.

Fixing a window size value of 100 sample, we performed a grid search for the optimal value of the sliding step length. Figure 2b shows a scatter plot representing the results of the search. As visible from the graph, searching in a range [10–40]% of the window size, we get similar results in fact all the accuracy values associated with the different sliding sizes resulted in a model accuracy of approximately 95.5%. We decided to pick a reasonably small value in the search



(a) *Window size* parameter tuning (b) *Sliding step length* parameter tuning

Fig. 2. Ball toy



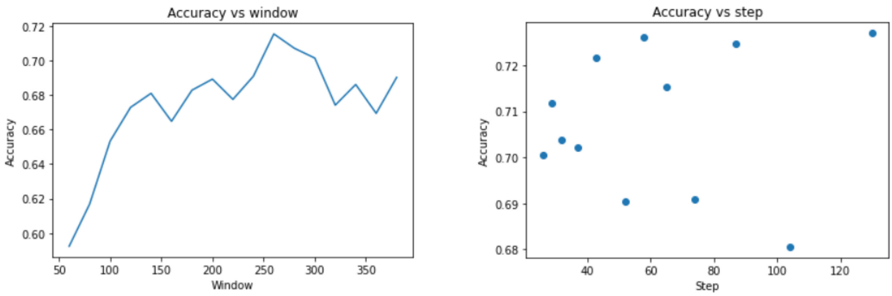
(a) *Window size* parameter tuning (b) *Sliding step length* parameter tuning

Fig. 3. Car toy

range (step size length = 20 samples, that is 20% of the window size), which allowed both to avoid overfitting and to have a consistent training dataset, at the same time. Summarizing, the procedure has identified an optimal window size of 100 samples (1 s) and an overlay step of 20 samples (0.2 s).

The same process has been performed with the **car** toy related data (Fig. 3a). In this case the identified optimal window size corresponds to 220 samples (2.2 s). Also in this case we plotted (Fig. 3b) the activity prediction accuracy versus the sliding step size: there is no clear trend in the plot, considering a search range of [20–100] samples. The related accuracy ranges from 93.1% to 93.9%: we have selected 50 samples as optimal step, because it is producing the highest accuracy score.

For what concerns the **elephant** toy, Fig. 4a shows the plot of accuracy vs window size. We have selected the optimal window size located on the global maximum of the range, more precisely at a window size of 260 (2.6 s) samples. As we can observe in Fig. 4b, the accuracy does not show a clear trend in correlation with the sliding step length, and it floats around the value 70.5%. A sliding step size of 60 samples (0.6 s) has been selected.



(a) *Window size* parameter tuning

(b) *Sliding step length* parameter tuning

Fig. 4. Elephant toy

In order to proceed with point 5 in the data workflow (*model training*) we calculated the 131-dimensional features vectors aggregating on the time windows defined above, to create the training set. The correlation between the obtained features has been calculated. No particular cases of high correlated features has been observed: the highest correlation value obtained is around 40%.

We proceed with the analysis training three well known classifiers: a *random forest* [10], an *ada boost* [3] and a *gradient boost* [3] classifier. For the models parameters tuning we performed a grid search over multiple parameters values ranges, in order to find the best parameters values. We adopted an 8-fold stratified cross validation to assess the performances of the trained model, averaging on the *accuracy* per each fold.

5 Results

5.1 Results Based on the In-Lab Collected Dataset

As stated above, the dataset collected in lab and described in Sect. 3 was divided into two sets: using the 80% for the training phase and the remaining 20% for testing. From the selected possible classifiers mentioned in Sect. 4, we selected the *Gradient Boost* to be the best performing one. The tables presented below show the performances of the classifier trained separately on each of the toy-related dataset. In all the three cases, the trained classification models show very good performances.

Ball. As shown in Table 4, the classification model trained on the ball related dataset has good performances. The values reported have been calculated over a test set where the activity classes are balanced (as it is on the training set), after the parameters tuning phase. As it is possible to observe in the classification report, the accuracy is quite high with an average value of 97% on the in-lab collected test set.

Table 4. Ball in-lab dataset based classification report

	Precision	Recall	F1-score	Support
Throw	1.00	0.99	0.99	774
Roll over	0.92	1.00	0.96	774
Shake	1.00	0.92	0.96	774
Turn	0.99	0.99	0.99	774
Accuracy			0.97	3096
Macro avg	0.98	0.97	0.97	3096
Weighed avg	0.98	0.97	0.97	3096

Elephant. Similarly, also the classifier trained over the elephant related dataset performs very well (Table 5): also in this case the test average accuracy is 97%.

Car. Is worth noting that the car-data based classifier is performing slightly worse compared to the previously presented ones (performances shown in Table 6). This is mainly due to the fact that the classifier shows difficulties in distinguishing between the activities *turning wheel* and *slide*. This is mainly due to the fact that this toy, as in its lab configuration, has a single sensor node located inside a back wheel. In the case of *turning wheel* activity, the kid is turning the wheels with one hand, while holding the car with the other hand. Meanwhile the *slide* activity is performed by make the car wheels rolling on the ground. The resulting signal sampled by the sensors show only slight differences between the

Table 5. Elephant in-lab dataset based classification report

	Precision	Recall	F1-score	Support
Let it fall	0.98	0.95	0.96	220
Overturn	0.96	0.95	0.96	220
Knock	0.96	0.99	0.98	220
Throw	0.98	0.99	0.98	220
Accuracy			0.97	880
Macro avg	0.97	0.97	0.97	880
Weighed avg	0.97	0.97	0.97	880

two actions. The solution adopted for this problem is to install two sensor nodes in the car, one in a front wheel and the other in a rear wheel. With this sensors configuration, we distinguish the turning wheel activity from the sliding activity (where both wheels roll at approximately the same rate): this configuration has been exploited in the real world use-case sampling sessions.

Table 6. Car in-lab dataset based classification report

	Precision	Recall	F1-score	Support
Slide	0.99	0.91	0.95	296
Overturn	0.98	0.93	0.95	296
Knock	0.94	0.98	0.96	296
Turning wheels	0.91	0.99	0.95	296
Accuracy			0.95	1184
Macro avg	0.95	0.95	0.95	1184
Weighed avg	0.95	0.95	0.95	1184

5.2 Validation of the In-Lab Data Base Classification Models

To asses the performances of the models trained over the in-lab collected data, it is important to perform a validation, which allows us to measure how well the proposed methodology is able to generalise: this is carried out exploiting an additional in-lab collected dataset, which the models have never seen before, and assessing the related performance. This validation step concerning in-lab data, has been carried out on data collected by two adult people (a man and a woman), which freely played with the AutoPlay toys, performing autonomously sequences of activities from the list of predefined ones, selected per each specific toy (see Sect. 3).

Table 7. Validation over in-lab collected data

Toy	Precision	Recall	F1-score	Accuracy
Car	0.84	0.71	0.70	0.71
Ball	0.96	0.96	0.96	0.96
Elephant	0.73	0.73	0.70	0.71

Table 7 shows the performances measured during this validation phase: it shows that the trained models achieve good classification performances both in terms of precision, recall, F1-score and accuracy. The *ball* related activity classification shows the best results, achieving 96% of accuracy (weighted average per activity). The *elephant* and *car* related activity classification experiences lower accuracy scores of 71%: however, given the hardness of the classification task, due to the peculiarity and noise of the activities, we could consider the achieved score to be acceptable. The main problem associated with the car related classification, as mentioned in the previous section, reseeds in the fact that it wrongly predict the 'slide' activity as 'turning wheel' activity. Regarding the elephant toy, on this validation phase, the trained model locates most of its prediction error for the activity 'Trow', which is interpreted erroneously as 'Knock': this is because a part of the 'Trow' activity includes a phase in which the elephant toy hits the ground, and in terms of forces acting on the toy and measured by the accelerometer these are similar to the ones measured during the 'Knock' activity. A dynamical window size approach could eventually mitigate this issue.

5.3 Validation on Real Use-Case Scenarios

In order to evaluate the performances of the proposed models on a real use case scenario involving children, we performed a second validation phase. This includes data collected from real world experiments involving small children (average target age is 2 years old). For the data collection, the AutoPlay complete toy set is provided to 3 children, which independently and autonomously play with them, without any request from or interaction with an adult person. The children are observed and data is collected through embedded sensors and cameras, as for the in-lab measurements described above. The complete data collected correspond to approximately 30 min of measurement. The measurement environment is shown in Fig. 5.

Fig. 5. Real use-case scenario: measurement environment

The results obtained by feeding the collected data to the models trained with the in-lab scenario data, show promising performances. The trained models are performing very well in recognizing rotation related activities (i.e., reaching 79% accuracy on ball 'roll-over' prediction, and 92% precision on car 'slide' prediction), confirming the validation step presented above. However, we notice that the trained models show difficulties with exploratory activities (i.e., car 'overturn' and ball 'shake' have a very low recall). Since the quantity and variety of child activity data which is feasible to collect in a real scenario will most likely be not sufficient for a traditional model training, we envision the possibility to apply *transfer learning* methodologies for the classification purposes in our future works.

6 Conclusion

In this paper we exploit current SoA activity recognition methodologies for wearable sensors, applying them on non-worn, non intrusive sensors. More specifically, the mentioned methodologies are applied on a smart toy set developed in the context of the AutoPlay project. We present activity prediction methodologies, trained over in-lab collected data, and capable to predict infant play activity from inertial data collected by sensor nodes embedded within the toys. The usage of non intrusive measurement methodology is beneficial in preserving the infant behaviour while playing, thus in gathering meaningful, reliable and objective data. This aspect is particularly important when the measurement is exploited for clinical observation purposes. Being able to make prediction on acquired data from non-worn, non-intrusive sensors is thus essential. In this paper we have produced an in-lab dataset using the AutoPlay toys, we annotated the activities with the aid of cameras and we trained activity prediction models for a limited set of three toys and four activities per each toy (identified as the most relevant infant activities). The results obtained are really promising: the achieved test accuracy on in-lab data ranges from 71% (for car and elephant toys) to 96% (for a ball toy). Therefore the proposed study demonstrates that the activity recognition applied to sensors data coming from toys embedded sensor nodes has great potentials and that the available technology is ready for such implementation.

Acknowledgements. We would like to acknowledge first of all Emmanuelle Rossini for the definition of the idea at the base of the AutoPlay approach. We acknowledge Dr. Gian Paolo Ramelli and his neuropediatric team of the EOC Hospital in Bellinzona for supporting us in the real use case scenario data collection. We acknowledge Franesca Faraci, Alessandro Puiatti, all the SUPSI DTI and DEASS collaborators of the AutoPlay team which contributed to the collection of the real use case scenarios data, Pepe Hiller and Hape Toys for the AutoPlay toys design and production. Last but not least, we would like to thank all the families and children which trust our project and which agreed to collaborate with us.

References

1. Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., Reyes Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pp. 437–442 (2013)
2. Antar, A.D., Ahmed, M., Ahad, M.A.R.: Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: a review. In: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision and Pattern Recognition (icIVPR), pp. 134–139. IEEE (2019)
3. Bahad, P., Saxena, P.: Study of adaboost and gradient boosting algorithms for predictive analytics. In: Singh Tomar, G., Chaudhari, N.S., Barbosa, J.L.V., Aghwariya M.K. (eds.) International Conference on Intelligent Computing and Smart Communication 2019. Algorithms for Intelligent Systems. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-0633-8_22
4. Faraci, F.D., et al.: Autoplay: a smart toys-kit for an objective analysis of children ludic behavior and development. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6. IEEE (2018)
5. Mehmood, A., Raza, A., Nadeem, A., Saeed, U.: Study of multi-classification of advanced daily life activities on shimmer sensor dataset. Int. J. Commun. Networks Inf. Secur. **8**(2), 86 (2016)
6. Papandrea, M., Giordano, S.: Location prediction and mobility modelling for enhanced localization solution. J. Ambient. Intell. Humaniz. Comput. **5**(3), 279–295 (2014)
7. Rivera, D., García, A., Alarcos, B., Velasco, J.R., Ortega, J.E., Martínez-Yelmo, I.: Smart toys designed for detecting developmental delays. Sensors **16**(11), 1953 (2016)
8. Sguazza, S., et al.: Sensor data synchronization in a IoT environment for infants motricity measurement. In: EAI International Conference on IoT Technologies for HealthCare, pp. 3–21. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-42029-1_1

9. Shimmer: Shimmer faqs. http://www.shimmersensing.com/support/wireless-sensor-networks-faqs/
10. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P.: Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. **43**(6), 1947–1958 (2003)
11. Westeyn, T.L., Abowd, G.D., Starner, T.E., Johnson, J.M., Presti, P.W., Weaver, K.A.: Monitoring children's developmental progress using augmented toys and activity recognition. Pers. Ubiquit. Comput. **16**(2), 169–191 (2012)