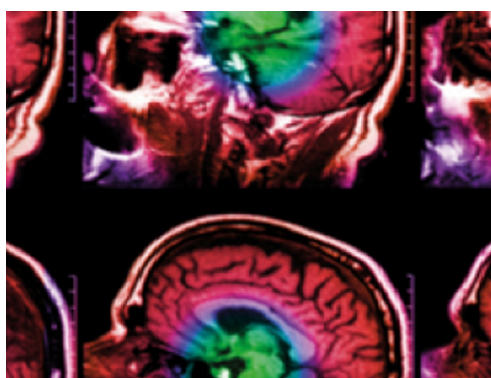


PAPER • OPEN ACCESS

Hidden Markov model-based activity recognition for toddlers

To cite this article: Mark V Albert *et al* 2020 *Physiol. Meas.* **41** 025003

View the [article online](#) for updates and enhancements.



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.

OPEN ACCESS

PAPER



Hidden Markov model-based activity recognition for toddlers

RECEIVED
23 September 2019REVISED
18 December 2019ACCEPTED FOR PUBLICATION
22 January 2020PUBLISHED
4 March 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Mark V Albert^{1,2,3,4,7} , Albert Sugianto³, Katherine Nickele⁵, Patricia Zavos⁵, Pinky Sindu³, Munazza Ali¹ and Soyang Kwon^{5,6} ¹ Department of Computer Science and Engineering, University of North Texas, Denton, TX, United States of America² Department of Biomedical Engineering, University of North Texas, Denton, TX, United States of America³ Department of Computer Science, Loyola University Chicago, Chicago, IL, United States of America⁴ Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, Chicago, IL, United States of America⁵ Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital, Chicago, IL, United States of America⁶ Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, United States of America⁷ Author to whom any correspondence should be addressed.E-mail: Mark.Albert@unt.edu

Keywords: activity recognition, HMM, toddler

Abstract

Objective: Physical activity has been shown to impact future health outcomes in adults, but little is known about the long-term impact of physical activity in toddlers. Accurately measuring the specific types and amounts of physical activity in toddlers will help us to understand, predict, and better affect their future health outcomes. Although activity recognition has been extensively developed for adults as well as older children, toddlers move in ways that are significantly different from older children, indicating the need for a more tailored approach. **Approach:** In this study, 22 toddlers wore Actigraph waist-worn accelerometers which recorded their movements during guided play. The toddlers were videotaped and their activities were later annotated for the following eight distinct activity classes: lying down, being carried, riding in a stroller, sitting, standing, running/walking, crawling, and climbing up/down. Accelerometer data were extracted in 2 s signal windows and paired with the activities the toddlers were performing during that time interval. **Main results:** A variety of classifiers were tuned to a validation set. A random forest classifier was found to achieve the highest accuracy of 63.8% in a test set. To improve the accuracy, a hidden Markov model (HMM) was applied by providing the predictions of the static classifiers as observations. The HMM was able to improve the accuracy to 64.8% with all five classifiers increasing the accuracy an average of 1.3% points (95% confidence interval = 0.7–1.9, $p < 0.01$). When the three most misclassified activities (sitting, standing, and riding in a stroller) were collapsed together, the accuracy increased to 79.3%. **Significance:** Further refinement of the toddler activity recognition classifier will enable more accurate measurements of toddler activity and improve future health outcomes of toddlers.

Introduction

Physical inactivity is known to contribute to a variety of negative health outcomes including obesity and diabetes (Jensen *et al* 2014). Studies conducted with children 5 years of age show that physically inactive children tend to follow the same physically inactive trajectory later in life (Janz *et al* 2009, Kwon *et al* 2015). Even children under the age of five have been shown to be inactive and it is important to understand the relationship between inactivity even earlier in life, such as during toddlerhood, to be able to influence that trajectory toward lifelong health.

With accurate assessment of toddler physical activity, the link to future health outcomes could be more clearly established. With improved data collection, community intervention can then be considered and individual health interventions can be better justified. Subjective reports of physical activity are poor compared to objectively measured data (Shephard 2003), and this has been observed when comparing survey assessments of activity from caregivers of young children to their measured activity (Noland *et al* 1990). Additionally, it is impractical

to have clinicians observe toddlers for extended periods of time. However, wearable devices can provide objective information about physical activity (Van Cauwenberghe *et al* 2011, Kate *et al* 2016). Still, estimating physical activities using wearable devices can be challenging for populations with different styles of mobility (de Almeida Mendes *et al* 2018, Kwon *et al* 2019), such that the movements of unique populations may not be identified correctly without taking their unique movements into consideration. For example, a model that trained on the physical activity patterns of non-Parkinson participants performed poorly when applied to participants with Parkinson's (60.3% accuracy). However, when a model was trained on the physical activity patterns of Parkinson's participants specifically, the model performance dramatically improved (92.2% accuracy) (Albert *et al* 2012a).

Machine learning, based on data from wearable devices, provides a straightforward way to tailor activity recognition models to particular populations. To date, this has enabled improved activity recognition accuracy for a variety of smaller clinical populations including Parkinson's patients (Albert *et al* 2012b), transfemoral amputees (Albert *et al* 2013, 2014), and stroke patients (O'Brien *et al* 2017). Toddlers require their own activity recognition models because they move differently from older children or adults, as they are in the developmental stage for upright movements such as walking, running, and jumping. In addition to tailoring activity recognition to the population, there are certain activities that are done only in specific populations. First, measuring activities for some distinct groups might be different, such as those with wheelchair use (Sok *et al* 2018), tremors (Albert and Kording 2011, Albert *et al* 2012a), and falls (Albert *et al* 2012b, Shawen *et al* 2017). In the case of toddlers, 'being carried', 'riding in a stroller', 'crawling', and 'climbing up/down' are common activities that would not be observed in most other populations, but are important in assessing the nature of toddler activity. Therefore, it is critical to acquire data from toddlers directly to train the systems that will be applied specifically to them.

Another factor in developing an activity recognition model is that data collected in a lab setting or under instruction is often less varied than when similar activities occur in a free-living setting. Previous activity recognition studies for pre-school age children have conducted structured or semi-structured activity trials in laboratory settings, providing specific instructions regarding the order and the length of activities to perform (Zhao *et al* 2013, Hagenbuchner *et al* 2015, Trost *et al* 2018). For example, in an activity recognition study involving adult participants with incomplete spinal cord injury, it was observed that accuracy dropped to 54.6% when testing a classifier on at-home activities in a lab setting, but was increased to 85.6% when the classifier was trained on data collected in an at-home setting (Albert *et al* 2017). A similar trend has been observed for toddler activity level prediction. When an activity classification (accelerometer count cut-point) algorithm developed for toddlers in a lab setting was applied to data collected during free play, the area under the receiver operating characteristic curve (ROC-AUC) validity ranged between 0.5 and 0.7 (Van Cauwenberghe *et al* 2011). As such, it is critical to use data collected in a natural setting rather than in a lab setting when developing an activity recognition model.

In addition, given the variability in toddler movements, the static window-based classifiers can be improved by incorporating the prior and later context of the classification. One tool to accomplish this is the hidden Markov model (HMM). HMMs combine uncertain observations over time to optimally infer a sequence of states—in this case, activities performed. HMMs have been shown to improve activity recognition in adult populations (Antos *et al* 2014, Sok *et al* 2018).

With the goal to conduct accurate assessment of toddler physical activity, the aims of this study were to develop machine learning classifiers for eight distinct activities performed by toddlers and to examine whether the performance of the classifiers can be improved using a HMM. To our knowledge, no published studies to date have trained a HMM model using toddler activity recognition. This study is one of the first studies to use a HMM to try to augment the estimates made by static window-based classifiers for toddler activity recognition.

Methods

Data was collected from 22 toddlers (12 females) aged between 13 to 35 months. Toddlers were recruited among the users of a private indoor child playroom located in Chicago. The two inclusion criteria for recruitment were age (13 to 35 months) and the ability to walk independently. Two of the original 24 toddlers were excluded from analysis due to errors in syncing their accelerometer data and the recorded video data. Written consent was obtained from their parents as approved by the Institutional Review Board (IRB) of Ann & Robert H. Lurie Children's Hospital of Chicago.

Data collection

Participants wore tri-axial ActiGraph wGT3X-BT (ActiGraph, Pensacola, Florida, USA) accelerometers on an elastic waist belt that was positioned around the waist. Participants had no issue of wearing the elastic belt. The x , y , and z axis of the ActiGraph accelerometer generally directed leftward, upward, and forward relative to the child, respectively. Data from waist-worn accelerometers were extracted using the ActiLife software and processed using custom Python scripts. The accelerometer sensor captured three-axial acceleration at a rate of 30 Hz. The physical activities of the toddlers were videotaped for later annotation. The annotations were originally

Table 1. The annotated toddler activities and the number of samples collected from each activity.

Activity	Number of extracted 2 s windows
Run/walk	933
Crawl	85
Climb	141
Stand	410
Sit	772
Lie down	128
Carried	160
Stroller	177
Totals	2806

made for 20 different activities. Three authors were involved in annotation and when an annotation was unclear, a majority vote was used to label the activity. In this way, the recording was annotated with a resolution of one second. Further details on the annotation process are available in previous related work (Kwon *et al* 2019). From the original 20 activities, a number of those activities were rarely observed, only performed for short periods of time, or were not sufficiently distinct from a clinical point of view to warrant separate classification. For that reason, only a representative eight activities were used consisting of lying down, being carried, riding in a stroller, sitting, standing, running/walking, crawling, and climbing up/down. Table 1 indicates the variation observed in the amounts of each movement. The average time annotated for each participant was 13 min (range: 6–21 min). Figure 1 shows sample windows of the accelerometer signals for each of the eight activities.

Data feature extraction

Accelerometer signal data was segmented into 2 s time windows to generate samples to train the classifiers. Time domain and frequency domain features were extracted from the segmented windows. Samples taken from the accelerometer were only used for training if the entirety of the window fell during the same activity.

A total of 76 statistical features were extracted from each 2 s window of the 3-axis accelerometer. Table 2 lists the features extracted from each signal including mean, standard deviation, max, min, and other standard signal processing measures. Time series signals included x , y , and z axes as well as a vector magnitude signal. Frequency signals were also generated from each of these axes, and magnitude signals were generated using the fast Fourier transform (FFT). These frequency signals are useful for quantifying the amount and frequency of periodic motion in the signal.

Hyperparameter tuning and testing

Subject-wise cross-validation was used to establish the efficacy of the models. In order to tune hyperparameters for each model, a grid search was performed using 10-fold cross-validation on the data from all-but-one subject. The parameter space searched for each model is shown in table 3. The hyperparameters which most often provided the highest accuracy on the 10-fold cross-validation are shown in bold.

HMM parameter settings

Next, a HMM was used to examine whether it can improve the accuracy of the window-based classifier done previously. Instead of working directly from a subset of features, the observations/emissions for the HMM consisted of the static window-based classifier outputs—probabilistic estimates from the classifiers when available. The emission probability model used was a Gaussian mixture model with eight different outputs for the eight different activity classes. The mean for each emission probability was the fraction of correctly identified windows from the static classifiers, while the variance was equal for all classes. This distribution reflects the uncertainty in the estimate of a given static classifier and is directly related to the confusion matrix of the given static classifier, such as the one shown in figure 2. The transition probability matrix for the HMM was constructed from the probability of transitions as observed for the collected window data (table 4). The same transition matrix was used for all participants.

Results

Table 5 presents the accuracies of the static window-based classifiers and HMM augmented classifiers. The highest accuracy of 63.8% was achieved with the random forest classifier. For all the classifiers, the percent improvement above the window-based classifier using the HMM was on average 1.3% points (0.7–1.9, 95% confidence interval [CI], $p < 0.0001$).

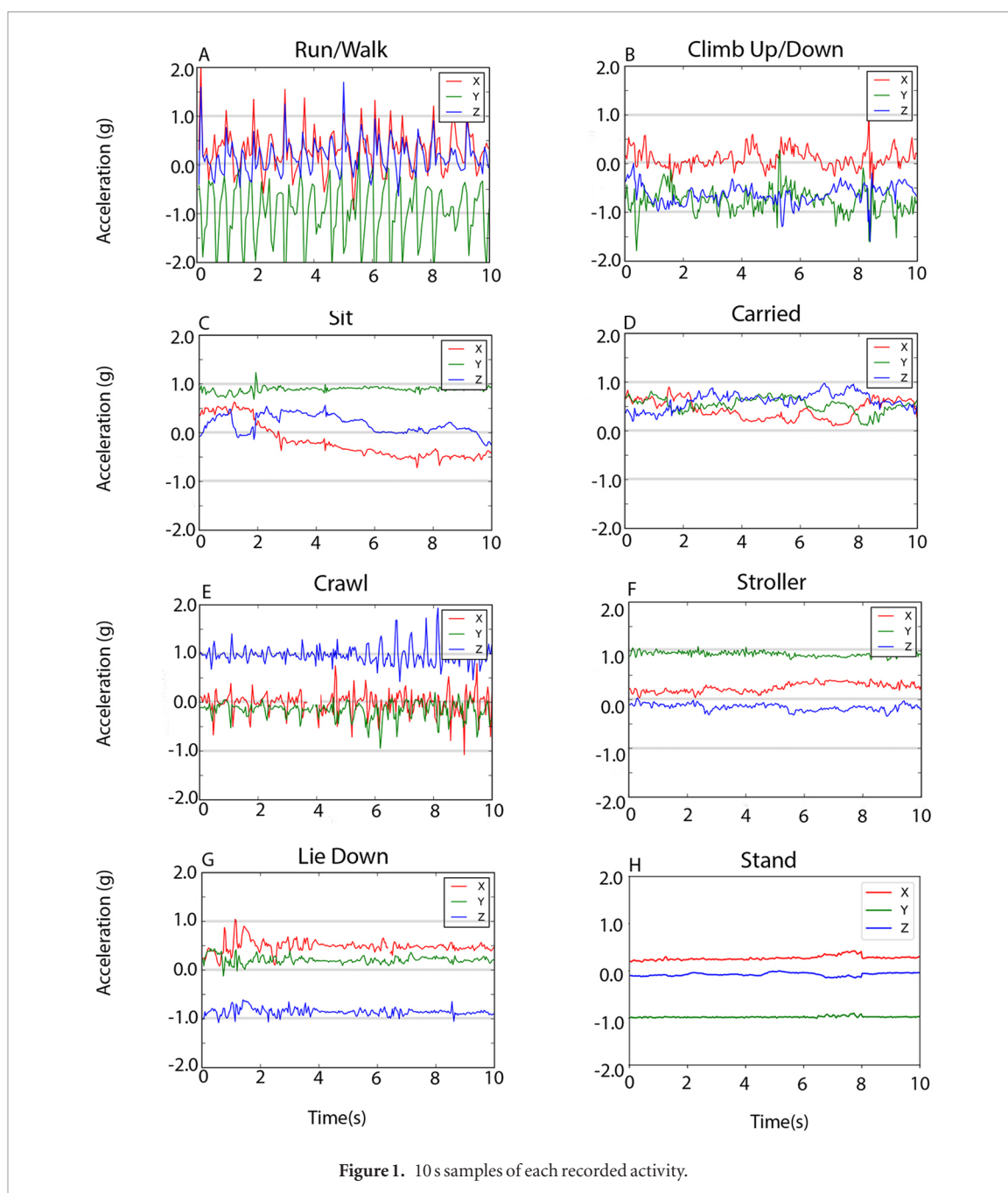


Figure 1. 10 s samples of each recorded activity.

Table 2. Features extracted from each 2 s window. Note that features were extracted for each axis directly (x, y, z), the magnitude of the axes, and FFT transformations of each time-domain signal.

Description	Features	Signal	Count
Moments	Mean, standard deviation, skew, kurtosis	$x, y, z, \sqrt{x^2 + y^2 + z^2}$	16
Quantiles	Minimum, median, maximum	$x, y, z, \sqrt{x^2 + y^2 + z^2}$	12
Cross correlation	Mean	xy, yz, xz	3
FFT moments	Mean, standard deviation	FFT of x, y, z	6
FFT quantiles	Minimum, median, maximum	FFT of x, y, z	9
FFT bins	10-bin histogram	FFT of x, y, z	30

To quantify the recognition of the random forest classifier for each of the 8 classes, we have provided the recall and precision values with a 95% CI for each class in table 6. To observe the nature of the types of misclassifications made, the confusion matrix is presented in figure 2. The confusion matrix compares the true negatives, false negatives, true positives, and false positives for the labeled and predicted activities. Additionally, samples that were mislabeled by the classifier are presented in figure 3 for visual observation of their similarity to the classes they match. Notably, the most confused activities match expectations for sensor similarity as shown in the example

Table 3. Hyperparameter grid of searched values and (bold) values most often chosen for testing.

Classifier	Hyperparameter	Values (bold—most selected)
K-nearest neighbors	Number of neighbors	1, 3, 5, 7, 9, 11, 13, 15
SVM	Regularization parameter	1×10^{-6} , 1×10^{-5} , ..., 1 , ..., 1×10^3 , 11×10^4
	Kernel coefficient	1×10^{-6} , 1×10^{-5} , ..., 1×10^{-1} , ..., 11×10^3 , 11×10^4
Logistic regression	Inverse of regularization strength	1×10^{-5} , 1×10^{-4} , ..., 1×10^{-1} , ..., 11×10^3 , 11×10^4
Decision tree	Min samples split	2, 4, 6, 8, 10
	Min samples leaf	1, 5, 10, 15 , 20
	Max depth	10 , 20, 30, 40, 50
Random forest	Number of estimators	10, 20, 30, 40, 50 , 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400

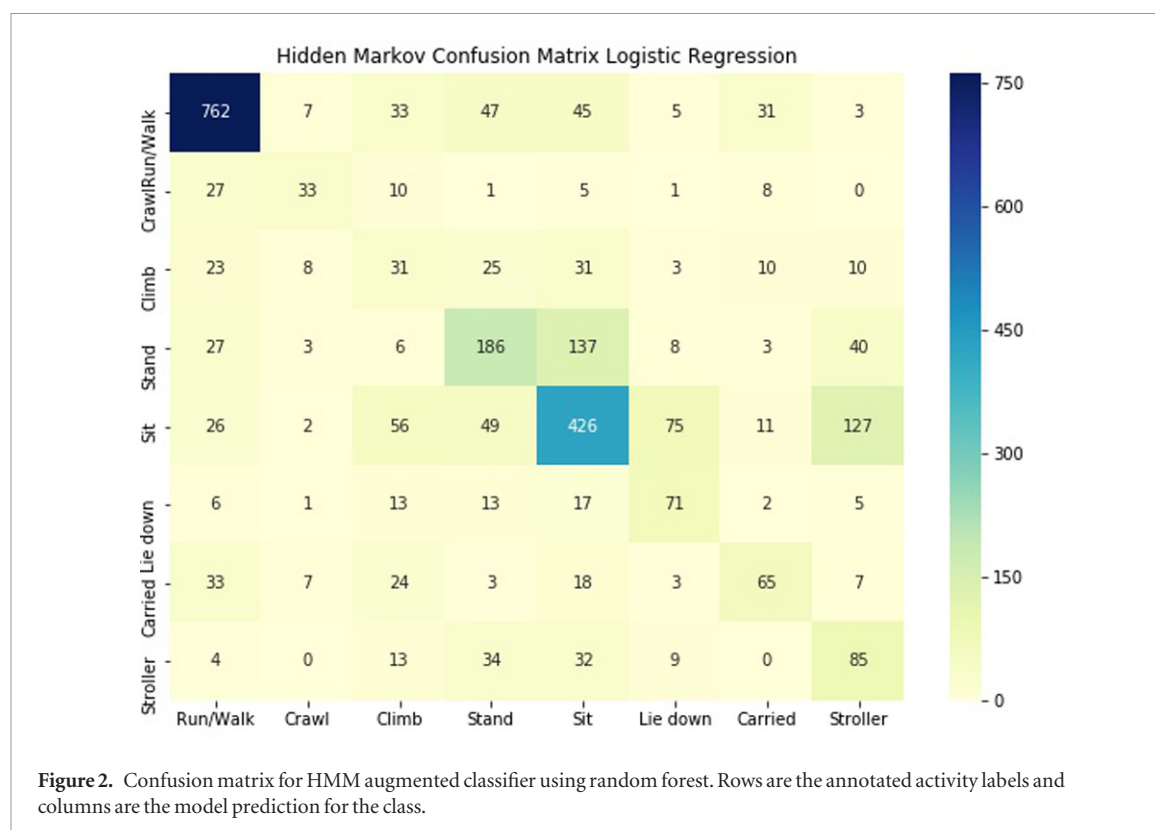


Table 4. The HMM transition probability matrix used as derived from the collected data. The source activity is the row and the destination activity is the column with the row sum equal to 1.

	R/W	Crawl	Climb	Stand	Sit	Lie	Carried	Stroller
Run/walk	0.782	0.012	0.025	0.115	0.036	0.008	0.018	0.004
Crawl	0.153	0.647	0.012	0.047	0.106	0.012	0.023	0.000
Climb	0.113	0.014	0.688	0.064	0.099	0.014	0.007	0.000
Stand	0.239	0.010	0.017	0.659	0.049	0.005	0.012	0.010
Sit	0.061	0.014	0.009	0.017	0.869	0.009	0.006	0.014
Lie down	0.086	0.008	0.000	0.039	0.047	0.805	0.016	0.000
Carried	0.081	0.006	0.031	0.006	0.031	0.038	0.794	0.013
Stroller	0.028	0.000	0.006	0.006	0.073	0.000	0.006	0.881

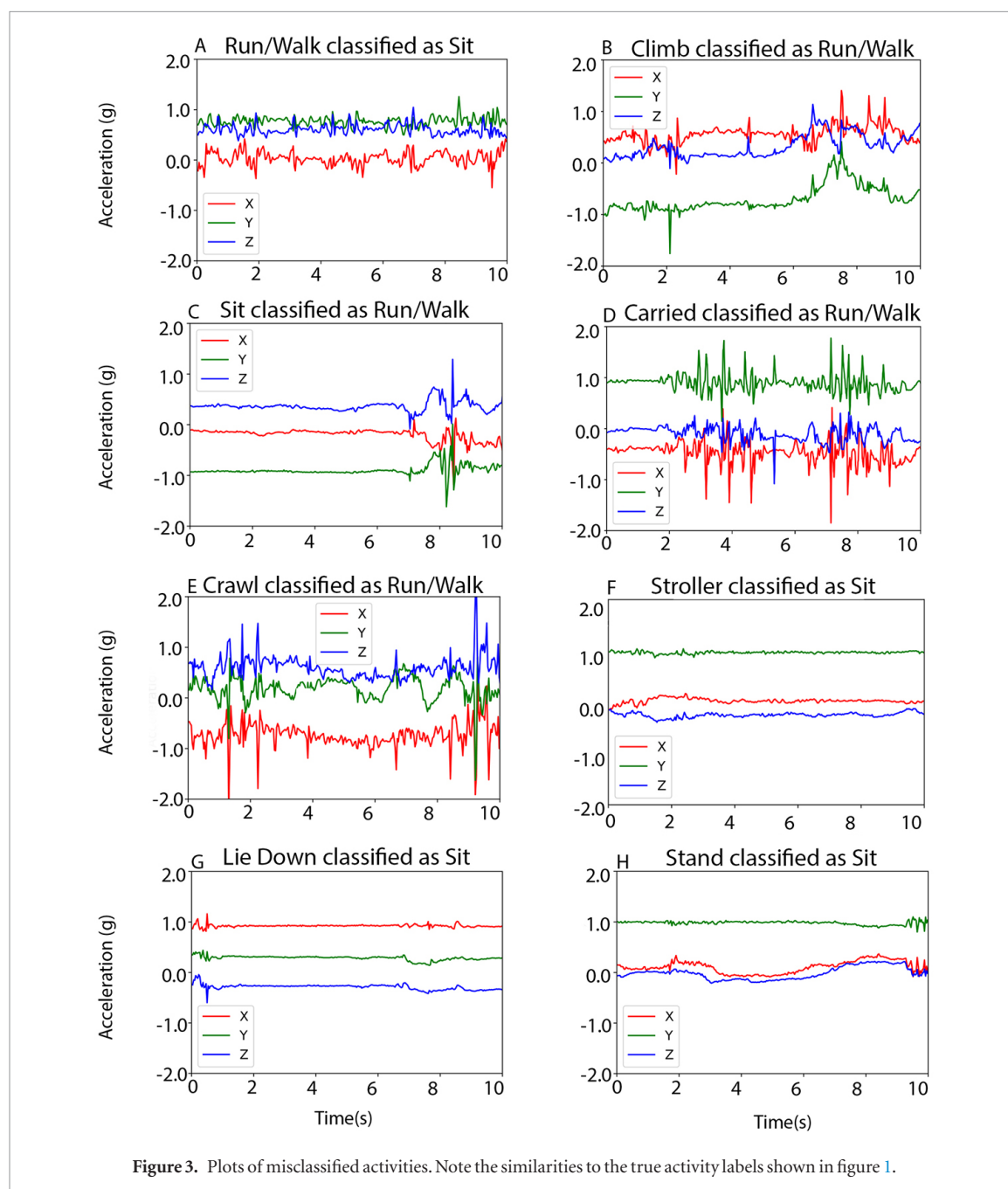
Table 5. Overall accuracy of the static window-based activity recognition classifier before and after HMM augmentation.

	Random forest (%)	SVM (%)	Logistic regression (%)	Decision tree (%)	K nearest (%)
Static window classifier	63.8	58.6	57.0	57.3	52.6
HMM augmented classifier	64.8	60.1	59.1	57.5	54.1
Improvement	1.0	1.5	2.1	0.2	1.5

Table 6. Recall, precision, and most common misclassification for each class using the random forest classifier with the HMM augmentation.

Activity	Recall (95% CI)	Precision (95% CI)	Most common misclassification (misclassification %)
Run/walk	80.0% (77.2–82.5)	88.2% (85.9–90.4)	Carried (7.3%)
Crawl	81.2% (71.2–88.8)	68.3% (58.3–77.2)	Climb (11.8%)
Climb	56.0% (47.4–64.4)	29.0% (23.7–34.8)	Crawl (10.6%)
Stand	45.4% (40.5–50.3)	49.2% (44.1–54.4)	Sit (32.7%)
Sit	66.6% (63.1–69.9)	66.9% (63.5–70.3)	Stand (12.0%)
Lie down	61.7% (52.7–70.2)	76.0% (66.6–83.8)	Climb (17.2%)
Carried	58.8% (50.7–66.5)	43.5% (36.8–50.4)	Run/walk (18.8%)
Stroller	28.2% (21.8–35.5)	41.0% (32.2–50.3)	Sit (44.1%)

Note: CI, confidence interval.



activity plots of figure 1. For example, 'sitting' and 'standing' are often confused with 'riding in a stroller' which have very similar sensor profiles when the child is stationary. Notably, 'being carried' is often confused with 'running/walking' primarily because when using waist-worn accelerometer signals it may be difficult to distinguish 'being carried' from the walking movement of a carrying adult.

Discussion

One of the challenging aspects of evaluating activity recognition systems is that the style and difficulty of tasks can vary widely. The eight activities chosen were representative of visually recognizable activities of toddlers during play, with less initial attention to the challenge presented in distinguishing these activities using wearable devices. For example, the accuracy for the eight assessed activities would increase from 64.8% to 79.3% if the three most confused passive activities were grouped together ('standing', 'sitting', and 'riding in a stroller'). In short, it is important in evaluating these results to put the context of the problem into account.

Machine learning has been used to tailor activity recognition in children (Zhao *et al* 2013, Nam and Park 2013, Trost *et al* 2014, Chowdhury *et al* 2017, Trost *et al* 2018). In a study by Trost and colleagues (Trost *et al* 2014), seven activities performed by young teenagers were classified using machine learning with an accuracy of nearly 90% using hip or wrist-worn sensors. These seven activities, lying down, sitting, standing, walking, running, as well as performing basketball and dance, although were succinctly described as similar because their recorded processed counts in the vertical axis were similar (Trost *et al* 2014), it is important to note that in this study there were specifically instructed movements associated with each activity. For example, sitting involved handwriting and playing a computer game while standing was composed of three tasks that involved upper body movements: throw and catch, a laundry task, and sweeping the floor. Classifying standing versus sitting with fixed, instructed activities in a laboratory setting is a more straightforward task than classifying these activities when the children are stationary or in more natural contexts as with our study.

There has been less research on machine learning-based activity recognition among children under age 5 years. Among those studies, another study by Trost (Trost *et al* 2018) measured the activities of preschoolers at age 3–6 years, collapsing 12 separate activities into five activity groups (sedentary, light activity games, moderate-to-vigorous games, walking, and running) using wrist and hip-worn sensors. Zhao *et al* (2013) also studied preschool age children and achieved a similar accuracy to Trost (Trost *et al* 2018) by separating activities into five classes (rest, quiet play, low active play, moderately active play, and very active play). Nam and Park (2013) studied infants and toddlers using a waist-worn accelerometer applying a wide variety of classifiers. Hagenbuchner *et al* (2015) used a deep learning ensemble network to classify 12 distinct instructed activities in a group of 3 to 6 year olds. Accuracies in activity recognition among preschool-aged children generally varied between 70% and 90%, but often these accuracies were achieved using fewer activities or controlled conditions and instructions for exhibiting unique movement patterns.

In the activity recognition literature, it is also well documented that movements in daily living are often more varied and difficult to track than lab-based and instructed movements (Kerr *et al* 2016, Bourke *et al* 2016, Albert *et al* 2017, Kerr *et al* 2017). Kerr *et al* (2017) notes that 'movement in the laboratory setting may not reflect free-living physical activity behavior because laboratory-based movements generally occur in sequences defined by the investigator'. Another study by Kerr (Kerr *et al* 2016) points to concerns associated with lab-acquired movements when working with participants who are obese or have comorbidities. Previous work in patient populations demonstrates that by training systems using data acquired from activities performed at home, where there is greater variability in movement styles, the accuracy of the models can be increased significantly (Albert *et al* 2017). For this reason we believe it is critical to not only acquire active movement data for validation, but to also use those movements to train recognition systems expected to function appropriately in natural settings, as was one of the aims of this study.

Our final aim was to demonstrate the improvement of window-based classifiers using an HMM. Improvement was demonstrated with a modest average gain in accuracy of 1.3% points. There are a number of advanced machine learning approaches that have been applied to improve upon traditional window-based predictive models, with HMMs consistently demonstrating improvements over traditional methods. Ellis *et al* (2016) demonstrated improvements using an HMM model compared to a cut-point based method to identify one of four activities (sitting, standing, walking/running, and riding) during one week of free play behavior, however this study was conducted among a population of overweight women. Pober *et al* (2006) also trained a HMM model for activity recognition to recognize the activities (walking, walking uphill, vacuuming, working at a computer) of adults and found a 10% point increase in recognition accuracy compared to quadratic discriminant analysis (QDA). Other advanced machine learning methods led to accurate predictions. Hagenbuchner (Hagenbuchner *et al* 2015) used a deep learning model and Chowdhury (Chowdhury *et al* 2017) used an ensemble method to perform activity classification for 12 different activities of children and adults. de Almeida Mendes (de Almeida Mendes *et al* 2018) reviewed several studies about activity recognition in adults as well as children and compared

their accuracy and R^2 values to study accelerometer usage in machine learning activity recognition. However, as stated before, to our knowledge, no published studies to date have trained a HMM model using toddler activity recognition. This study is one of the first studies to use a HMM to augment the estimates made by static window-based classifiers for toddler activity recognition.

There is also a variety of signal window sizes used in activity recognition work on children. These range from 1–10 s (Trost *et al* 2014) to 11–30 s or more (Van Cauwenberghe *et al* 2011, Trost *et al* 2012, 2018). Generally, if classification is done on isolated windows, larger window sizes are preferred. However, smaller window sizes are needed for the resolution necessary to capture certain activities. We chose 2 s windows as a tradeoff between a window large enough to capture enough signal for window-based estimates but short enough to pick up quickly-varying movements. Notably, identified activities can vary widely even within a single identified window. This can be observed in figure 1, where it can be seen that although the activity is identified as one class there are portions of the window which indicate other distinct movements.

The most challenging aspect in activity recognition design is determining which activities are clinically relevant and weighing that against the difficulty in reliably identifying those activities and considering how separable they are given the available movement data. For example, some activities are inherently difficult to accurately classify due to the limitations of the accelerometer signal which can be a source of error. For example, ‘sitting’ and ‘standing’ may be dramatically different types of activities when considering the context of those activities. However, it is difficult to distinguish the two using the accelerometer signals if no motion is present as the orientations are the same. Similarly, ‘sitting’ compared to ‘riding in a stroller’ may be difficult to distinguish. However, we chose to present these difficult classes separately given the clinical relevance. Although the physiological activity of the child is similar, the engagement of the caregiver is significantly different which may provide useful information for future studies. There is an expectation that with improved recognition methods that take temporal context into account, the overall amount of each of these activities could be estimated.

Another source of error in this study may have come from errors in the annotation file extracted from acquired video of the toddler activity. The annotation file is used as the ground truth for labeling the toddler activity; and to achieve the best accuracy it is critical to have proper signal window labels. Additionally, although precautions were taken to sync the start times of the video annotations and accelerometer signals, any discrepancy of more than one second could lead to many mislabeled windows. Additionally, in future studies, the activities selected for annotation could more closely match the specific types of movements rather than the holistic activities we have described here.

Conclusion

We developed machine learning classifiers to classify eight distinct activities for toddlers and then improved the performance using a HMM-activity recognition model with a relatively low overall accuracy (64.8%). However, when the three most misclassified activities (sitting, standing, and riding in a stroller) are collapsed together the accuracy increases to 79.3%. Although the improvements brought by augmenting the static classifiers with an HMM were modest, only 1.3% on average across classifiers, we would expect this to improve with more data where static classifiers reach a limit of what can be observed in a single clip. As such, future studies that include more data are warranted. In summary, given the inherent challenges in tracking toddler movements, especially during free play, machine learning techniques provide a means to create and validate activity recognition systems. With properly validated toddler activity recognition, clinical researchers would be able to explore the link between toddler physical activity and future health outcomes. These findings could support the need for early intervention and lead to improved lifelong health outcomes.

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

ORCID iDs

Mark V Albert  <https://orcid.org/0000-0003-3977-2895>

Soyang Kwon  <https://orcid.org/0000-0002-9359-704X>

References

- Albert MV and Kording K P 2011 Determining posture from physiological tremor *Exp. Brain Res.* **215** 247–55
- Albert MV *et al* 2013 Monitoring functional capability of individuals with lower limb amputations using mobile phones *PLoS One* **8** e65340

- Albert M V et al 2014 Monitoring daily function in persons with transfemoral amputations using a commercial activity monitor: a feasibility study *PM R* **6** 1120–7
- Albert M V, Azeze Y, Courtois M and Jayaraman A 2017 In-lab versus at-home activity recognition in ambulatory subjects with incomplete spinal cord injury *J. NeuroEng. Rehabil.* **14** 10
- Albert M V, Kording K, Herrmann M and Jayaraman A 2012a Fall classification by machine learning using mobile phones *PLoS One* **7** e36556
- Albert M V, Toledo S, Shapiro M and Kording K 2012b Using mobile phones for activity recognition in Parkinson's patients *Frontiers Neurol.* **3**
- Antos S A, Albert M V and Kording K P 2014 Hand, belt, pocket or bag: Practical activity tracking with mobile phones *J. Neurosci. Methods* **231** 22–30
- Bourke A K et al 2016 Video analysis validation of a real-time physical activity detection algorithm based on a single waist mounted tri-axial accelerometer sensor *Conf. Proc. IEEE Engineering in Medicine Biology and Society* pp 4881–4
- Chowdhury A K, Tjondronegoro D, Chandran V and Trost S G 2017 Ensemble methods for classification of physical activities from wrist accelerometry *Med. Sci. Sports Exerc.* **49** 1965–73
- de Almeida Mendes M et al 2018 Calibration of raw accelerometer data to measure physical activity: a systematic review *Gait Posture* **61** 98–110
- Ellis K et al 2016 Hip and wrist accelerometer algorithms for free-living behavior classification *Med. Sci. Sports Exerc.* **48** 933–40
- Hagenbuchner M et al 2015 Prediction of activity type in preschool children using machine learning techniques *J. Sci. Med. Sport* **18** 426–31
- Janz K F et al 2009 Sustained effect of early physical activity on body fat mass in older children *Am. J. Prev. Med.* **37** 35–40
- Jensen M D et al 2014 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society *J. Am. Coll. Cardiol.* **63** 2985–3023
- Kate R J, Swartz A M, Welch W A and Strath S J 2016 Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data *Physiol. Meas.* **37** 360–79
- Kerr J et al 2016 Objective assessment of physical activity: classifiers for public health *Med. Sci. Sports Exerc.* **48** 951–7
- Kerr J et al 2017 Comparison of accelerometry methods for estimating physical activity *Med. Sci. Sports Exerc.* **49** 617–24
- Kwon S et al 2015 Active lifestyle in childhood and adolescence prevents obesity development in young adulthood *Obesity* **23** 2462–9
- Kwon S et al 2019 Hip and wrist-worn accelerometer data analysis for toddler activities *Int. J. Environ. Res. Public Health* **16** (<https://doi.org/10.3390/ijerph16142598>)
- Nam Y and Park J W 2013 Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor *IEEE J. Biomed. Health Inform.* **17** 420–6
- Noland M, Danner F, DeWalt K, McFadden M and Kotchen J M 1990 The measurement of physical activity in young children *Res. Q. Exerc. Sport* **61** 146–53
- O'Brien M K et al 2017 Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting *J. Med. Internet Res.* **19** e184
- Pober D M, Staudenmayer J, Raphael C and Freedson P S 2006 Development of novel techniques to classify physical activity mode using accelerometers *Med. Sci. Sports Exerc.* **38** 1626–34
- Shawen N et al 2017 Fall detection in individuals with lower limb amputations using mobile phones: machine learning enhances robustness for real-world applications *JMIR Mhealth Uhealth* **5** e151
- Shephard R J 2003 Limits to the measurement of habitual physical activity by questionnaires *Br. J. Sports Med.* **37** 197–206
- Sok P et al 2018 Activity recognition for incomplete spinal cord injury subjects using hidden Markov models *IEEE Sens. J.* **18** 6369–74
- Trost S G et al 2012 Identification and validity of accelerometer cut-points for toddlers *Obesity* **20** 2317–9
- Trost S G et al 2018 Sensor-enabled activity class recognition in preschoolers: hip versus wrist data *Med. Sci. Sports Exerc.* **50** 634–41
- Trost S G, Zheng Y and Wong W-K 2014 Machine learning for activity recognition: hip versus wrist data *Physiol. Meas.* **35** 2183–9
- Van Cauwenberghe E, Gubbels J, De Bourdeaudhuij I and Cardon G 2011 Feasibility and validity of accelerometer measurements to assess physical activity in toddlers *Int. J. Behav. Nutr. Phys. Act.* **8** 67
- Zhao W et al 2013 Support vector machines classifiers of physical activities in preschoolers *Physiol. Rep.* **1** e00006