Identification of Developmental Delay in Infants using Wearable Sensors: Full-Day Leg Movement Statistical Feature Analysis

Mohammad Saeed Abrishami[1], Luciano Nocera[2], Melissa Mert[3], Ivan A. Trujillo-Priego[4], Sanjay Purushotham[2], Cyrus Shahabi[2] and Beth A. Smith[4]

Abstract—This paper examines how features extracted from full-day data recorded by wearable sensors are able to differentiate between infants with typical development and those with or at risk for developmental delays. Wearable sensors were used to collect full-day ($8 - 13$ hours) leg movement data from infants with typical development ($n = 12$) and infants at risk for developmental delay ($n = 24$). At 24 months, at-risk infants were assessed as having good ($n = 10$) or poor ($n = 9$) developmental outcomes. With this limited size dataset, our statistical analysis indicated that accelerometer features collected earlier in infancy differentiated between at-risk infants with poor and good outcomes at 24 months, as well as infants with typical development. This paper also tested how these features performed on a subset of the data for which the infant movement was known, i.e., 5-minute intervals more representative of clinical observations. Our results on this limited dataset indicated that features for full-day data showed more group differences than similar features for the 5-minute intervals, supporting the usefulness of full-day movement monitoring.

I. INTRODUCTION

Mobility assessment is an important clinical tool used to identify individuals with or at risk for mobility impairments, and to optimize and individualize intervention.

Current mobility analysis typically relies on brief observations performed by a trained clinician using a clinical rating scale. One scale commonly used for assessing infant mobility is the Alberta Infant Motor Score (AIMS) [1]. Assessments relying on scales have several potential shortcomings: (1) infants may behave differently when examined in different settings (e.g., home vs. clinic), (2) the observation period might be insufficient for the infant to demonstrate his or her full repertoire of skills, (3) trained healthcare professionals are needed, and (4) the evaluation is based on subjective visual observations. Wearable sensors have been proposed as a method to overcome these shortcomings [2].

Specifically, with the advancement and pervasiveness of wearable sensors (e.g., Apple Watch, Fitbit), it is now possible to continuously collect full-day data from individuals. Thus, there is an unprecedented opportunity to augment a patient's clinical visits with these longitudinal datasets. Compared to traditional monitoring that is most often done at the clinic, monitoring with wearable technology can be less intrusive and less expensive, and allows collecting data about overall activity and health for a longer duration of time. Importantly, wearable devices can allow observation of infants in their natural environment.

While wearable devices have been applied successfully in the case of identification of cerebral palsy (CP) [3], [4], [5], they have not been validated when applied to the identification of infants more broadly at risk of developmental delay. This paper presents a study in which wearable sensors were used to record leg movements over a full day from infants broadly at risk of neuromotor developmental delay.

We analyzed how features extracted from the raw sensor data (accelerometer and gyroscope) were useful in differentiating between infants at risk of developmental delay and infants with typical development. We further examined how the features discriminated between poor or good development outcomes for the at-risk category assessed at 24 months. Finally, to evaluate the usefulness of the full-day data, the performance of these features was compared to that of a smaller subset of the data, corresponding to a period of 5 minutes. This subset of data was chosen for its temporal correspondence to typical mobility assessments done in a controlled clinical environment.

The study design, data collection, data pre-processing, and features used in the analysis are presented in Section II. The statistical analysis methodology is described in Section III. The results are presented in Section IV. We further discuss our results and related works in V. Finally, we provide conclusions and discuss future work in Section VI.

II. DATASET

This section explains the study design, describes the dataset, and elaborates upon the feature extraction.

A. Study Design

In this study*, full-day leg movement data were collected from a group of 12 infants with **typical development (TD)** and 24 infants **at risk for developmental delay (AR)**†. At risk for developmental delay is defined based on population-based

[1]Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA (e-mail: abri442usc.edu)

[2]Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA (e-mail: {nocera, spurusho, shahabi}@usc.edu)

[3]Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA (e-mail: mkoc@usc.edu)

[4]Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA 90033, USA (e-mail: {itrujill, beth.smith}@usc.edu)
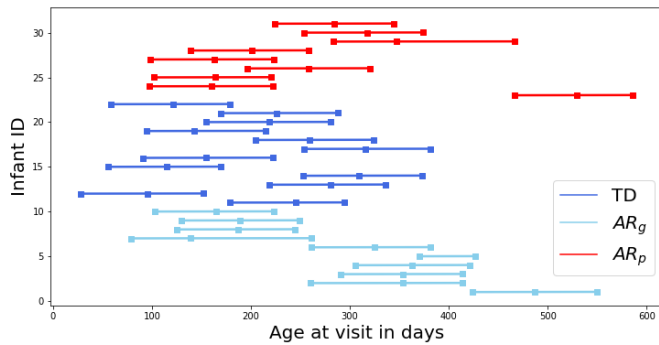
Fig. 1: Infant age at time of visit: Connected lines indicate each infant, while squares represent each visit. TD = typical development. AR_g/AR_p = at risk with good/poor outcome at 24 months.

Fig. 2: Infant wearing sensors on the front of each ankle.

criteria including pre-term birth and complications at or after birth. The criteria used to define at risk for developmental delay can be found, in full, in the reference [8]. Infants with *TD* were from singleton, full-term pregnancies with scores above the 5th percentile on Alberta Infant Motor Scale (AIMS) [1], [9].

During the study, infants were monitored for three complete days, at two-month intervals. Age at first visit was $1-8$ months for *TD* infants and $2-15$ months (adjusted for prematurity) for *AR* infants as shown in Figure 1. The infants were visited at their homes each morning, and wearable sensors were placed on each of their ankles. Wearables were attached with Velcro to a knee sock and covered by a second sock or custom leg-warmers with a pocket to hold them in place as shown in Figure 2. Families were encouraged to go about their typical daily activities. Infants wore the sensors until bedtime, resulting in about $8-13$ hours of data.

Most visits included a 5-minute video of the infant's spontaneous movement. In this period, infants were awake, alert, and content. Infants below the age of 7 months were recorded while they were in a supine position. Infants aged 7 months and older were recorded while supported in a standing position (held at the trunk), in order to prevent them from rolling or crawling away during the recording. These recorded videos were later used by an expert to annotate the movements of the infants to provide a ground truth, confirming the accuracy of the movement detection algorithm introduced in Subsection II-C. The data from this short period, referred to as **controlled environment** data, bears similarity to the typical clinical measurements. It should be noted that not all infant visits included a controlled environment recording. In total, 120 measures were included in this specific measurement (60 including data from both left and right leg sensors, with each pair collected in different visits).

In follow-up, the parent or guardian was asked if the child had any diagnoses at 24 months of age. This information was used to differentiate at-risk infant outcomes. A diagnosis of developmental delay was labeled as *poor* (**AR$_p$**), while no diagnosis was labeled as *good* (**AR$_g$**).

The characteristics of infants who participated in the study are summarized in Table I. Four infants whose families could not be reached for follow-up were excluded, as was one who passed away.

B. Sensor Data

We used APDM Opal wearable sensors [10] (comprised of 3D-accelerometer, 3D-gyroscope, and 3D-magnetometer). APDM sensors are wireless, small ($48.4mm \times 36.1mm \times 12mm$) and lightweight ($22g$) and are thus well suited for this infant study. The acceleration range is $\pm6g$, and measurements are reported with 14-bits resolution.

Recordings were made at 20Hz. Recorded data were stored on the internal memory of each individual sensor. The data of both left and right sensors were actively synchronized throughout the recording and were downloaded at the end of each visit. The video recordings were acquired at 30 frames per second (fps) and later synchronized with wearable data. During each visit, infants' anthropometric measurements (weight, length, and head circumference) were measured, while motor development status was quantified using the AIMS [1].

C. Pre-processing and Feature Extraction

For Opal sensors it was reported [11] that pre-processing the acceleration to remove the effect of gravity introduces further noise. This was independently verified for the purposes of this study. Accordingly, the raw acceleration signal was utilized.

To extract single leg movements of infants, we used an algorithm described in [6]. The algorithm was validated using the video recordings of infants on the same dataset used for this study. The algorithm distinguishes separate leg movements when a leg pauses or changes direction. For each single movement extracted, the features introduced in [7] were computed: duration of a movement, peak acceleration and average acceleration during a movement. In our statistical analysis we included the mean value of the feature computed as the mean of the daily feature average.

III. ANALYSIS

This section analyzes the correlation between left and right leg sensor data and introduces the statistical methods utilized.

TABLE I: Baseline characteristics of samples by developmental group, (N=31, TD: Typical development, AR_g/AR_p: At risk of developmental delay with good/poor outcome)

Group	N	Age	Corrected Age [a] <days>	Male	Length <cm>	Weight <kg>	Head <cm>	AIMS [b]	Developmental Stage [c]		
									C	SNC	NS
TD	12	162 (74.5-211.5)	162 (74.5-211.5)	4	60.5 (59.3-67.5)	6.4 (5.4-8.3)	40.5 (38.5-45.3)	19 (8.0-27.5)	0	2 (16.7%)	10 (83.3%)
AR_g	10	260.5 (125.0-305.0)	205.5 (103.0-229.0)	5	66 (57.0-72.7)	6.6 (5.5-8.5)	42.3 (38.0-44.0)	29.5 (9.0-32.0)	3 (30.0%)	2 (20.0%)	5 (50.0%)
AR_p	9	196 (102.0-254.0)	102 (88.0-154.0)	6	61 (61.0-63.5)	6.4 (5.9-7.3)	40.3 (40.0-41.2)	11 (7.0-15.0)	0	0	9 (100.0%)
p-value		0.1	0.45		0.83	>0.99	0.96	0.09	0.04		

[a]Post-term age in days, to account for premature birth
[b]Alberta Infant Motor Scale (AIMS) raw total score
[c]C=crawling; SNC=sitting but not yet crawling; NS=not yet sitting

A. Correlation between left and right legs' data

Since data were collected for both legs, Pearson correlation coefficients and their 95% confidence intervals (CI) were calculated in order to assess the degree of similarity for the salient features extracted from sensors of the left and right legs. A high correlation for any two methods designed to measure the same property might in itself suggest that a widespread sample has been chosen. A high correlation does not necessarily imply that there is good agreement between the two methods; consequently, in addition to Pearson-correlation analysis, Bland-Altman plots were generated to provide a visual representation of measurement agreement or bias between data from the two legs.

In this analysis, p-value (the significance level) implied the probability of the hypothesis that the correlation was due to chance. Results of this analysis are shown in Table II. Significant positive Pearson correlations were observed between right and left leg data for all outcome measures (all $p < 0.001$). This indicated that the results were not due to chance. As there was no evidence of any strong bias, both legs were considered for inclusion in the mixed model.

As an example, the scatter and Bland-Altman plots for mean duration are shown in Figure 3.

B. Statistical Methods

Sample characteristics at baseline (the first visit of every infant) were summarized and compared by developmental group. Continuous variables were expressed as median and interquartile range (IQR) and were analyzed by a Wilcoxon rank-sum test. Categorical variables were expressed as proportions and were analyzed by Fisher's exact test. Based on the results of section III-A, the data from both the left and right legs of the infants were used in the mixed model analysis.

TABLE II: Degree of similarity between right and left legs

Feature	N[a]	r[b]	CI[c]	p-value
Duration		0.89	$0.83 \leq \rho \leq 0.92$	$p \leq 0.001$
Peak Acceleration	91	0.93	$0.90 \leq \rho \leq 0.96$	$p \leq 0.001$
Mean Acceleration		0.92	$0.88 \leq \rho \leq 0.95$	$p \leq 0.001$

[a]Total number of samples
[b]The strength of the relationship
[c]95% Confidence Interval

Linear mixed effects models were used to assess the effects of developmental group on right and left leg movement duration, peak acceleration, and mean acceleration. Developmental group and age were included as fixed effects. Due to the variability of infant ages on-study, and the irregularity in the time interludes between visit waves for each infant, corrected age (defined as post-term age in days, to account for premature birth) was used as the time metric and modeled as a continuous random effect by right or left leg nested within infant. Significant effects were assessed using Tukey post-hoc comparisons.

Although several forms of covariance structure were tested, the variance components (VC) covariance structure was chosen due to the limited number of parameters available in our small sample. Since the rate of change of the outcome measurements across age can be expected to vary between infants in different developmental groups, a group $\times$ corrected age interaction was tested and included in the final model only if the term was significant at $p-value < 0.10$.

A $p-value < 0.10$ was considered significant for the present analysis given our limited sample size. All statistical analyses were performed using SAS version 9.4, (SAS Institute, Inc., Cary, NC).

C. Controlled Environment Data vs. Full-Day Data

The full-day data were obtained in an uncontrolled setting containing unknown factors. Data might contain activity generated by outside sources, such as a parent picking up the infant. Moreover, the mood of the infant can be variant over the period of a full day. Thus, a second analysis employed controlled environment data capture for a period of 5 minutes; movement capture in the controlled environment was constrained to activity exclusive to the child when he or she was alert and content. Yet, determining whether group differences could be inferred from noisy, full-day data was a principal goal of this study, as this kind of data capture represents typical activity and is environmentally valid. This investigation did not aim to identify or distinguish among the possible sources of noise in the full-day data. The same analysis procedures were done for the controlled environment data, as they were for Section III-B.

TABLE III: Mean Duration, Peak Acceleration, and Mean Acceleration by Developmental Group ($N = 31$, TD: Typical development, AR_g/AR_p: At risk of developmental delay with good/poor outcome)

Outcome [a]	Developmental Group			Group Effect	Age Effect	Group × Age Interaction
	TD	AR_g	AR_p			
Mean Outcome						
Duration	5.49 (0.05) [b]	5.27 (0.06)	5.22 (0.06) [c]	0.008	0.69	0.005
Peak Acceleration	13.50 (0.17)	13.22 (0.19)	12.78 (0.20) [c]	0.03	<0.001	NS
Mean Acceleration	10.56 (0.05)	10.58 (0.06)	10.39 (0.06) [bc]	0.04	<0.001	NS
Mean Rate of Change [d]						
Duration	0.01 (0.02)	-0.05 (0.02)	0.03 (0.01)	-	-	-

[a]Data represent least squares means and standard error in the format of [LS-means (SE)].
[b]Significantly different from AR_g group at Tukey-adjusted $p < 0.10$.
[c]Significantly different from TD group at Tukey-adjusted $p < 0.10$.
[d]Data represent mean rate of change (SE) for every one month (30 day) increase in infant age.

IV. RESULTS

A total of 31 infants, across an average of three visits, contributed 182 measures of right and left leg sensor data. Infants in different developmental groups were comparable across all baseline characteristics except developmental stage ($p = 0.04$;Table I). Developmental groups did not differ by
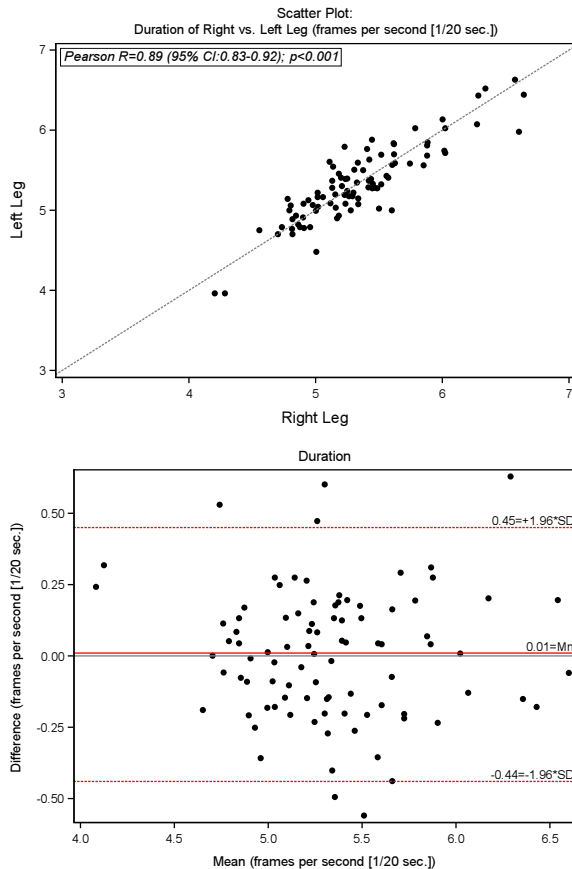

Fig. 3: Top: scatter plot of the mean movement duration of the left and right legs for the full day. Bottom: Bland-Altman plot. The diagrams illustrate how the measurements of the left and right leg sensor data were highly correlated. As there was no evidence of strong bias, both legs were included in the mixed model analysis.

median age or corrected age, even after the at-risk group was split into good and poor developmental outcome groups. To confirm, the association between developmental stage and corrected age was tested by a Wilcoxon rank sum test and found to be significant ($p = 0.004$).

For the mixed-effects models, a significant group effect was observed for all three outcomes (duration $p = 0.008$, peak acceleration $p = 0.03$, mean acceleration $p = 0.04$; Table III), indicating that the mean duration, peak acceleration, and mean acceleration of right and left leg movements all varied by group (TD, AR_p, AR_g). Typically developing infants' mean duration of right and left leg movements [LS-mean (SE): 5.59(0.05)] were significantly higher than both AR_p [LS-mean (SE): 5.22(0.06)] and AR_g [LS-mean (SE): 5.27(0.06)] infants at $p < 0.10$. Although the mean duration did not vary with age, the effect of age on duration varied by developmental group (interaction $p = 0.005$). AR_p infants' rate of duration change increased by 0.03(0.01) units per month of infant age, while TD infants increased by only 0.01(0.02) units and AR_g infants decreased by 0.05(0.02) units. Because there were no significant group differences in the mean rate of peak and mean acceleration progression, the interaction terms were dropped from the relevant models to avoid over-parameterizing. The mean peak acceleration of right and left leg movements was significantly lower in AR_p infants compared to TD infants [LS-means (SE): 12.78(0.20) and 13.50(0.17) for AR_p and TD infants, respectively]. For mean acceleration, significant differences were observed between AR_p infants [LS-mean (SE): 10.39(0.06)] and both TD and AR_g infants [LS-mean (SE): 10.56(0.05) and 10.58(0.06), respectively].

The use of corrected ages in our models accounted for variability in infant ages between developmental group and across different visit waves. Since our age adjustments incorporated both a statistically and biologically significant association between age and developmental stage, observed differences in developmental stage at baseline were not considered to be of concern.

An additional analysis was performed on the data obtained from the 5-minute controlled condition. Mixed models were run as described in the methods section. A total of 26 infants contributing 120 measures of right and left leg sensor data were included in this analysis. A significant group effect was

TABLE IV: Mean Duration, Peak Acceleration, and Mean Acceleration by Developmental Group Under Controlled Conditions ($N = 31$, TD: Typical development, AR_g/AR_p: At risk of developmental delay with good/poor outcome)

Outcome [a]	Developmental Group			Group Effect	Age Effect	Group × Age Interaction
	TD	AR_g	AR_p			
Mean Outcome						
Duration	5.50 (0.14)	5.28 (0.14)	5.63 (0.12)	0.18	0.41	NS
Peak Acceleration	13.61 (0.36) [b]	14.77 (0.35)	13.33 (0.30) [c]	0.02	0.22	NS
Mean Acceleration	10.53 (0.14)	10.78 (0.12)	10.33 (0.11) [c]	0.23	<0.001	0.07
Mean Rate of Change [d]						
Mean Acceleration	0.11 (0.04)	0.10 (0.03)	0.01 (0.03)	-	-	-

[a]Data represent least squares means and standard error in the format of [LS-means (SE)].
[b]Significantly different from AR_g group at Tukey-adjusted $p < 0.10$.
[c]Significantly different from TD group at Tukey-adjusted $p < 0.10$.
[d]Data represent mean rate of change (SE) for every one month (30 day) increase in infant age.

observed for peak acceleration ($p = 0.04$), but not for duration or mean acceleration as shown in Table IV. The mean peak acceleration of right and left leg movements was significantly lower in AR_p [LS-mean (SE): $13.33(0.30)$] and TD infants [LS-mean (SE): $13.61(0.36)$] compared to AR_g [LS-mean (SE): $14.77(0.35)$] infants at $p < 0.10$. Although there were no differences by group, the effect of age by group varied for mean acceleration; the average rate of mean acceleration of right and left leg movements was lower in AR_p infants [LS-means (SE): $0.01(0.03)$ units per month] compared to TD [LS-means (SE): $0.11(0.04)$ units per month] and AR_g infants [LS-means (SE): $0.10(0.03)$ units per month]. Due to the small sample size, results should be interpreted with caution.

V. Discussion

Wearables and nearables are a class of devices that have been used primarily for activity tracking and fitness monitoring. Fueled by recent advances in electronics miniaturization and soaring affordability, the newfound ubiquity of such devices coincides with enhancements in sensor capacity, and accuracy as well. A new arena arises, both for consumers in terms of visualizing their everyday behavior, and for researchers, who gain access to anonymous, environmentally valid motion data from millions of users. Numerous scholars have used these devices to advance their research interests.

Ravi et al. in [12] used data retrieved from accelerometers to recognize eight motion activities (standing, walking, running, climbing up stairs, climbing down stairs, sit-ups, vacuuming and brushing teeth). Brezmes et al. in [13] used accelerometer data from a mobile phone to recognize activities such as walking, climbing up and down, sitting up and down, and falling. Case et al. in [14] compared the accuracy of direct observation of step counts to that of smartphone applications and wearable devices. In [15] accelerometer data were used for de-ambulatory activity recognition using machine learning approaches. Bianchi in [16] surveyed recent developments in consumer and clinical devices for sleep, and these sensors have generated a wave of interest from clinicians for application in a wide range of diseases, such as Parkinson's disease [17], epilepsy [17], stroke [17], sleep disorders [16], and cardiac disorders [18], [19]. In these studies, wearable data were shown to be useful, even if noisy and containing inaccuracies, e.g., fitness bands generate relatively clean but potentially inaccurate activity data (step counts, activity level, calories burned) and noisy but accurate raw sensor data (accelerometer, gyroscope). Most wearable applications rely on processed activity data, yet for some applications such as infant mobility monitoring, adult activity models are not appropriate and features extracted from the raw sensor signals must be used.

Characteristics of spontaneous movements in infants at risk for CP have been studied and described in detail [20]. This assessment methodology relies on visual observation by expert clinicians. However, there is an emerging field of research using alternate methods such as motion capture cameras and wearable devices to automate this process and objectively measure infant movement [21]. Meinecke et al. in [3] used a motion capture system to record infants at high risk for CP and with TD for 15 minutes. From the recorded movement data they extracted 53 features, previously introduced by [20], that allowed them to differentiate between groups. In [22] the movements of ten pre-term infants in a Neonatal Intensive Care Unit were recorded for one hour using both accelerometers and video equipment. A physical therapist annotated the pre-defined abnormal movements by reviewing the videos. Using accelerometer data for the detection of these abnormal movements of interest showed promising results. To address the limitations of motion capture systems, Heinze et al. in [4] used four accelerometers on hands and feet to develop a high accuracy model for early prediction of CP.

In [23] Hadders-Algra illustrated that *variation* of movement behavior is the key factor for identifying children with, or at risk of, a developmental motor disorder. Their study also proposes that capturing this variation is more likely in longer periods of data capture.

In order to quantify infant movement behavior across a full day, Smith et al. [6] developed an algorithm to identify single leg movements in infants. They defined the start and end of each movement using acceleration and angular velocity thresholds. A new movement was identified each time the infant's leg paused or changed direction. They validated their algorithm using video recordings of infants' spontaneous movements and reported a sensitivity of 92%. Next, Trujillo-Priego et al. in [7] analyzed the kinematic characteristics of each identified movement. They calculated the duration of each movement, and also calculated the average and peak magnitude of the total acceleration during a given movement.

It has recently been proposed [24], [25] that to further advance the field of infant mobility assessment, new technologies must sample development for a minimum of 24 hour periods, so that the effects of circadian rhythms, behavioral context, environmental stimuli, mood and motivation, etc., may be taken into account. This study starts to do so, using full-day wearable recordings of infant leg movements. Participants were infants with TD and AR, and infants AR were retrospectively classified based on 24-month neuromotor outcomes. Pre-defined features were extracted from full-day accelerometer data and indicated that acceleration features differentiated between at-risk infants with poor developmental outcomes, at-risk infants with good developmental outcomes, and infants with typical development. Short period, controlled environment data sets collected over 5 minutes did not provide as much differentiation. Our results support the use of full-day wearable sensor data for early identification of developmental delay in infants.

VI. CONCLUSION AND FUTURE WORK

This study is unique and important as it analyzed full-day accelerometer data for infants, showing that simple features measured earlier in infancy can differentiate between infants at-risk of developmental delay who demonstrate poor or good outcomes at 24 months, and infants with typical development. Furthermore, our findings support the usefulness of wearable sensor data collected over long periods in an uncontrolled environment.

The limited number of samples in our dataset, as well as the broad ranges in age and developmental stage serve as some critical shortcomings within this investigation, thus the results should be used with caution. Further work is needed to validate our results on a larger dataset and to investigate features that can better model the characteristics of infant movement (we only analyzed features that mapped full-day data into single scalars). Ultimately, these features can be used to build diagnostic tools for the early identification of developmental delay in infants and for objective measurement of intervention outcomes.

ACKNOWLEDGEMENT

REFERENCES

[1] M. Piper and J. Darrah, *Motor Assessment of the Developing Infant*. Saunders Philadelphia, 1994, vol. 1, p. 176.

[2] M. Fan, D. Gravem, D. M. Cooper, and D. J. Patterson, "Augmenting gesture recognition with erlang-cox models to identify neurological disorders in premature babies," *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, p. 411, 2012.

[3] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, and C. Disselhorst-Klug, "Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy," *Human Movement Science*, vol. 25, no. 2, pp. 125–144, 2006.

[4] F. Heinze, N. Breitbach-Faller, T. Schmitz-Rode, and C. Disselhorst-Klug, "Movement analysis by accelerometry of newborns for the early detection of movement disorders due to infantile cerebral palsy," *IFMBE Proceedings*, vol. 25, no. 9, pp. 24–27, 2009.

[5] S. Ohgi, S. Morita, L. KK, and C. Mizuike, "Time series analysis of spontaneous upper-extremity movements of premature infants with brain injuries...including commentary by Fetters L and Sholz JP," *Physical Therapy*, vol. 88, no. 9, pp. 1022–1038, 2008.

[6] B. A. Smith, I. A. Trujillo-Priego, C. J. Lane, J. M. Finley, and F. B. Horak, "Daily quantity of infant leg movement: Wearable sensor algorithm and relationship to walking onset," *Sensors (Switzerland)*, vol. 15, no. 8, pp. 19 006–19 020, 2015.

[7] I. A. Trujillo-Priego and B. A. Smith, "Kinematic characteristics of infant leg movements produced across a full day," *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 4, 2017.

[8] California Department of Health Care Services. (Jan.) High risk infant follow up. Accessed: February, 2018. [Online]. Available: http://www.dhcs.ca.gov/services/ccs/Pages/HRIF.aspx#medicalcriteria

[9] J. Darrah, M. Piper, and M. Watt, "Assessment of gross motor skills of at-risk infants: predictive validity of the alberta infant motor scale," *Developmental Medicine and Child Neurology*, vol. 40, pp. 485 – 491, 1998.

[10] "Apdm wearable technologies," http://www.apdm.com/, accessed: September 1, 2017.

[11] "Using orientation estimates to convert from sensor frame to earth frame of refernce," http://support.apdm.com/hc/en-us/articles/214504186-Using-orientation-estimates-to-convert-from-sensor-frame-to-Earth-frame-of-refernce, accessed: 2017-09-08.

[12] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Aaai*, vol. 5, no. 2005, 2005, pp. 1541–1546.

[13] T. Brezmes, J.-L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone," in *International Work-Conference on Artificial Neural Networks*. Springer, 2009, pp. 796–799.

[14] M. A. Case, H. A. Burwick, K. G. Volpp, and M. S. Patel, "Accuracy of smartphone applications and wearable devices for tracking physical activity data," *Jama*, vol. 313, no. 6, pp. 625–626, 2015.

[15] M. Nguyen, L. Fan, and C. Shahabi, "Activity recognition using wrist-worn sensors for human performance evaluation," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, 2015, pp. 164–169.

[16] M. T. Bianchi, "Sleep devices: wearables and nearables, informational and interventional, consumer and clinical," *Metabolism: clinical and experimental*, 2017.

[17] D. Johansson, K. Malmgren, and M. A. Murphy, "Wearable sensors for clinical applications in epilepsy, parkinson's disease, and stroke: a mixed-methods systematic review," *Journal of neurology*, pp. 1–13, 2018.

[18] J. M. Pevnick, K. Birkeland, R. Zimmer, Y. Elad, and I. Kedan, "Wearable technology for cardiology: An update and framework for the future," *Trends in cardiovascular medicine*, 2017.

[19] M. Lang, "Beyond fitbit: A critical appraisal of optical heart rate monitoring wearables and apps, their current limitations and legal implications," *Alb. LJ Sci. & Tech.*, vol. 28, p. 39, 2017.

[20] H. F. R. Prechtl, "State of the art of a new functional assessment of the young nervous system. An early predictor of cerebral palsy," *Early Human Development*, vol. 50, no. 1, pp. 1–11, 1997.

[21] C. Marcroft, A. Khan, N. D. Embleton, M. Trenell, and T. Pl??tz, "Movement recognition technology as a method of assessing spontaneous general movements in high risk infants," *Frontiers in Neurology*, vol. 6, no. JAN, p. 284, 2015.

[22] D. Gravem, M. Singh, C. Chen, J. Rich, J. Vaughan, K. Goldberg, F. Waffarn, P. Chou, D. M. Cooper, D. Reinkensmeyer, and D. J. Patterson, "Assessment of Infant Movement With a Compact Wireless Accelerometer System," *Journal of Medical Devices*, vol. 6, no. 2, p. 021013, 2012.

[23] M. Hadders-Algra, "Variation and variability: key words in human motor development." *Physical therapy*, vol. 90, no. 12, pp. 1823–37, 2010.

[24] B. A. Smith, D. L. Vanderbilt, B. Applequist, and A. Kyvelidou, "Sample entropy identifies differences in spontaneous leg movement behavior between infants with typical development and infants at risk of developmental delay," *Technologies (Basel)*, vol. 5, no. 3, p. 55, Sep 2017, 29114479[pmid].

[25] K. E. Adolph and S. R. Robinson, "Sampling development," *Journal of Cognition and Development*, vol. 12, no. 4, pp. 411–423, Jan 2011, 22140355[pmid].