An Algorithmic Approach for Detecting Neuromotor Developmental Disabilities in Infants from Wearable Sensor Data

Maria Despoina Siampou*, Luciano Nocera*, Jinseok Oh‡, Beth A. Smith†‡, Cyrus Shahabi*
*Dept. of Computer Science, University of Southern California, Los Angeles, United States
†Dept. of Pediatrics, University of Southern California, Los Angeles, United States
‡Div. of Developmental-Behavioral Pediatrics, Children's Hospital Los Angeles, Los Angeles, United States

***Abstract*—The inherent challenges in recruiting human subjects, particularly infants, often hinder the acquisition of sufficiently large datasets for health research, thereby limiting the applicability of conventional machine-learning (ML) approaches. In this study, we analyze full-day motion recordings from two groups: typically developing infants (N = 12) and infants at risk for developmental disabilities (N = 24), further divided into those with good (N = 10) and poor (N = 9) developmental outcomes at 24 months. The goal is to differentiate at-risk (AR) infants from those with typical development (TD) and predict outcomes for the at-risk category using wearable data. Due to its limited size, previous studies on this dataset, employing statistical and machine learning methods, raise reliability concerns. To address this, we introduce a novel algorithmic approach to extract meaningful patterns, referred to as *M*otifs, from the raw signals. The abundance of Motifs serves as highly informative indicators, enabling effective differentiation between the groups. Evaluation on this limited-size dataset demonstrates the effectiveness of Motifs in distinguishing AR from TD infants and predicting future outcomes for the at-risk category.**

***Index Terms*—infant, neuromotor developmental disabilities, accelerometer data, motif discovery, matrix profile**

I. INTRODUCTION

The widespread integration of wearable devices into healthcare has facilitated the continuous monitoring of patients' vital signs and activities, presenting a promising alternative to periodic clinical visits. This approach finds applications across various domains [1]–[6], including the identification of neurological deficits in infants by measuring high-quality spontaneous movement patterns [7]–[9]. However, recruiting human subjects, particularly infants, is often challenging, resulting in datasets that may not be sufficiently large. This limitation restricts the applicability of traditional ML approaches as the next step, to analyze the collected data.

Typically, infant motor development status analysis relies on brief observations conducted by trained clinicians using clinical rating scales like the Alberta Infant Motor Score (AIMS) [10]. Such methods exhibit several limitations, including rater biases, short visit times, and the possibility of infants exhibiting unexpected behaviors in unfamiliar environments. Studying infant movements through wearable sensor data poses its own inherited challenges. Commercial wearable devices like fitness bands, primarily engineered for adult use, lack the specificity to capture infant movements, especially considering the remarkable movement variance exhibited during infancy. Additionally, obtaining large datasets is complicated by the need for guardian consent and the allocation of resources (financial, trained personnel, time) necessary to collect human subjects' data from infants. To address the first challenge, previous research has primarily focused on extracting and characterizing meaningful features from the raw signals, followed by statistical analysis [11] or supervised learning [12]. Yet, the limited number of study participants raises concerns about the reliability of the reported results.

In this work, we overcome these challenges by introducing an algorithmic approach that involves extracting meaningful patterns, referred to as *M*otifs. Our Motif-based approach offers several advantages over these traditional feature engineering-based methods. First, by identifying distinct repetitive patterns in our raw accelerometer signals, we effectively mitigate the noisiness of the recorded data while capturing higher-level information. Second, the abundance of Motifs forms a repository of interpretable patterns that can aid clinicians in identifying specific movement patterns within developmental groups. For instance, the Motif corresponding to an infant in a swing is interpretable and serves as a recognizable pattern to distinguish from infant movements effectively. This contrasts with features like acceleration norm or peak, which are challenging to filter out due to their fine-grained nature. Lastly, our results account for subtle variations that would otherwise go unnoticed. We show how our approach does not need any training to yield significant differences across developmental groups, thus making it ideal for healthcare applications where obtaining a large number of labeled data is difficult. By employing a simple time-series nearest-neighbor approach we reveal significant differences across the Motifs of typically developing (TD) and at-risk (AR) infants, as well as between infants with *poor* (AR_p) and *good* (AR_g) developmental outcomes within the at-risk category. To the best of our knowledge, this is the first attempt to use Motifs to compare distinctive patterns in infants with the explicit goal of predicting developmental disabilities.

II. METHODS

A. Participants & Data Collection

Full-day leg movement data was collected from two groups: 12 infants with typical development and 24 at risk for developmental disabilities. The dataset for infants with typical development was initially introduced in [13], [14]. APDM Opal wearable sensors were placed on the participants' ankles, measuring tri-axial acceleration and angular velocity at a sampling rate of 20 Hz. The participants were monitored for three days, with two-month intervals between each day, and wore the sensors until bedtime, resulting in approximately 8 to 13 hours of recorded data. Follow-up communication occurred when each infant reached 24 months old to determine their developmental status. Infants with *poor* developmental outcomes were identified as those receiving ongoing therapy services, while infants with *good* outcomes were those not receiving physical or occupational therapy. To that extent, this dataset includes 36 observations of TD infant leg movements and 70 AR infant observations[1], while follow-up developmental status information was obtained for 19 out of 24 AR participants, resulting in 55 AR infant movement recordings for analysis.

B. Data Preprocessing

To account for variations in sensor placement orientation among infants, we preprocessed the raw tri-axial signals. The magnitude of the acceleration vector for each recording was computed using the formula:

$$Accel_{mag} = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

where x, y, and z are the three directional components of acceleration. This was previously employed and validated in [13], [14] to detect leg movement occurrences using thresholding. Moreover, each signal underwent the following filtering process. First, a 3rd-order low-pass Butterworth filter was applied at the Nyquist frequency of 1 Hz to diminish high-frequency components, followed by signal downsampling to 1 Hz to accelerate computations.

C. Scalable Time series Anytime Matrix Profile (STAMP)

We aim to extract recurring patterns from the collected data, called *Motifs* [15]. To achieve that, we utilized the Scalable Time series Anytime Matrix Profile (STAMP) algorithm [16], [17]. Given a recording T with n data points and a desired pattern length m, STAMP returns the matrix profile P and the matrix profile index I, which correspond to the distances and indices of the best match (or nearest neighbor) for every subsequence S_i of size m in T. This is achieved by sliding a size window m over T and computing the pairwise euclidean distances between S_i and every other subsequence S_j. It is worth noting that the algorithm considers trivial matches, such as self-matches, by applying an exclusion zone of size $m/2$ before and after the location of each subsequence. This ensures accurate matching results and avoids redundant matches.

D. Time Series Motif Extraction

Having the profile P and the corresponding indexes I of each of our recordings, our pipeline moves towards identifying "meaningful" patterns. Here, we declare a motif meaningful when its distance to its best match is below a certain threshold. We define this threshold as follows:

[1]The data of the third visit of two AR infants were excluded since they were found to walk independently at the time.

$$th_i = \max((\mu(D_i) - 2\sigma(D_i)), \min(D_i)) \tag{2}$$

This indicates that each subsequence S_i becomes a candidate motif if it has at least one match with a distance less or equal to th_i. Note that, th_i varies for each subsequence. We have deliberately made that choice since infant movements exhibit varying noise levels and variability, and thus, using an adaptive threshold provides a more flexible solution for detecting patterns in our data.

Extracting the candidate motifs becomes straightforward with each entry in P and I corresponding to a subsequence in T. To that extent, we perform a single scan of the two vectors and keep the entries that satisfy our constraint. The remaining entries in I represent the starting points of the candidate motifs. Subsequently, we rank them in ascending order based on their distance to their best match and select the top k. If the number of candidate motifs is less than k, all are selected. Finally, we group the selected patterns of each right and left leg recording, resulting in a set of at most $2k$ Motifs for each reported session. In the next sections, we refer to this set as a *motif group*.

III. RESULTS & DISCUSSION

A. Experimental Setup

We present our results utilizing the pipeline presented in Section II on our collected infant leg movement recordings. We choose to extract patterns of different sizes (m) ranging from 10 to 50 seconds, to account for the wide range and variability of movements that infants normally exhibit. Finally, we experimentally set the number of extracted patterns per leg recording to $k = 4$. We evaluate our patterns for two distinct scenarios, presented below.

B. Typically developed vs at-risk infant patterns

Our first goal is to determine whether the extracted Motifs can indicate an infant at risk of developmental disabilities. To that extent, we employ the time-series k-nearest-neighbor classifier (kNN) and set the number of neighbors to 1 and the metric distance to *euclidean distance*. This approach allows us to investigate whether the motifs extracted from each infant are similar within their respective developmental groups while exhibiting differences across the groups. For evaluation, we adopt the Leave-One-Out Cross-Validation (LOOCV) technique and conduct two distinct experiments. In the first one, we evaluate the patterns of each recording independently, while for the second one, we group the Motifs across the three visits, treating the patterns of each infant as a single group. This arrangement allows us to investigate whether considering long-term observations would increase the chances of an AR infant being detected. For this scenario, our dataset consists of 106 recordings, comprising 36 TD and 70 AR infant observations.

Classification Results. The results, summarized in the left part of Table I, showcase the accuracy and recall of the classification model across different motif sizes. The model effectively can distinguish patterns of various sizes between the two classes, notably achieving 92% accuracy for $m = 50$. This highlights a clear dissimilarity in terms of the distance between the patterns of the two classes. The scores are

TABLE I: Prediction Accuracy and (Recall) for identifying *(i)* TD/AR (left) and *(ii)* AR_g/AR_p (right) infants using different motif sizes. The Motifs were evaluated individually for each recording and combined across visits for each infant.

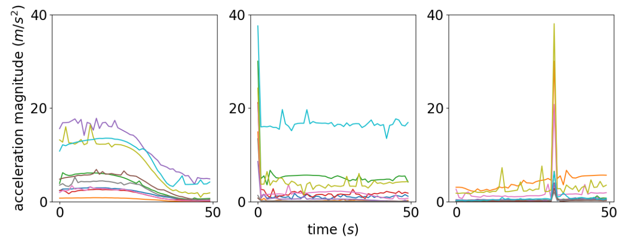
Class	**TD / AR**					**AR$_g$ / AR$_p$**				
motif size	m=10	m=20	m=30	m=40	m=50	m=10	m=20	m=30	m=40	m=50
Individual	0.78 (0.96)	0.82 (0.97)	0.86 (0.97)	0.89 (1.0)	*0.92 (1.0)*	0.53 (0.67)	0.60 (0.67)	0.63 (0.67)	0.62 (0.67)	*0.64 (0.74)*
Across Visits	0.81 (0.96)	0.86 (0.96)	0.86 (1.0)	0.94 (1.0)	**0.97 (1.0)**	0.58 (0.67)	0.58 (0.78)	0.63 (0.67)	0.68 (0.78)	**0.74 (0.89)**

further improved when the patterns are aggregated across visits, reaching 97% accuracy for $m = 50$. This suggests that extracting Motifs from data collected over multiple sessions is the most effective approach to assess an infant's developmental stage correctly. One possible explanation for this observation is that incorporating multiple sessions reflects the difference in the rate of infant development between the two groups. However, even when gathering such comprehensive data is not feasible, our results demonstrate that Motifs can still serve as valuable indicators to detect at-risk infants at an early stage. Lastly, the high recall achieved by the model ($\geq 96\%$) implies the absence of false negatives, highlighting the Motifs' efficacy as indicators for identifying at-risk infants. This is particularly important for our goal to support the healthy development of infants, by enabling early interventions.

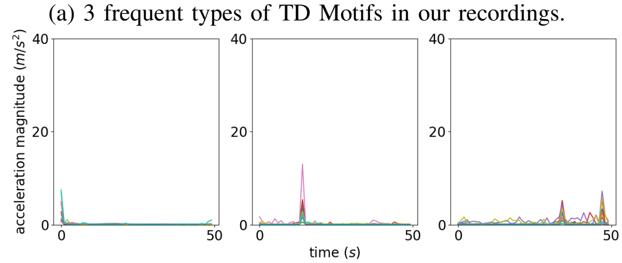
Motif Interpretability. We further explore the distinctions in Motifs between the two groups, as illustrated in Figure 1, showcasing the three most frequent Motif groups identified in our analysis. Notably, Motifs associated with typically developing (TD) infants demonstrate higher acceleration peaks than those of at-risk (AR) infants. To explore deeper, we extracted features from the Motifs of each class and conducted a statistical analysis. Specifically, we calculated the average *minimum*, *mean*, and *peak* of Motifs' acceleration, as well as their average *repetition* in the recordings from which they were extracted. Employing the Mann-Whitney U Test to assess the statistical significance of class differences, Table II presents the results for a pattern size of $m = 50$. Motifs of TD infants exhibit, on average, higher min (p<0.001), peak (p=0.012), and mean (p<0.001) acceleration compared to AR Motifs. This aligns with a previous report [11], supporting that TD infants display movements with greater intensity, particularly in peak acceleration per movement, compared to their AR peers. Although Motifs in our analysis are not identical to movements in the previous study, the finding suggests a consistent trend. We also note that the average repetition of TD patterns is lower compared to AR peers (p<0.001). This implies that TD infants tend to exhibit greater variability in their leg movements, resulting in fewer repeating patterns within the recorded data. This finding aligns with the results of an exploratory study that compared the variability of acceleration magnitude between TD and AR using sample entropy [18].

C. Predicting outcomes for at-risk infants

Our second objective is more ambitious, exploring whether Motifs can indicate the future developmental outcome of infants at 24 months of age. For this task, our dataset consists of 55 recordings, including 27 AR$_p$ and 28 AR$_g$ infants. We



(a) 3 frequent types of TD Motifs in our recordings.



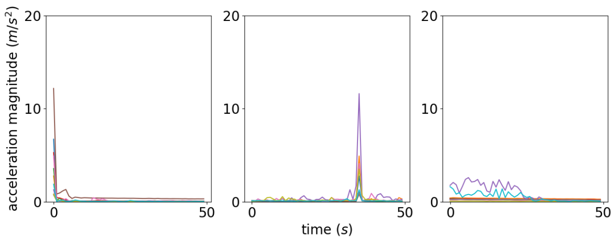
(b) 3 frequent types of AR Motifs in our recordings.

Fig. 1: Plots illustrate frequent types of (a) TD and (b) AR Motifs. Instances may vary slightly due to movement or noise, but share a qualitative similarity.

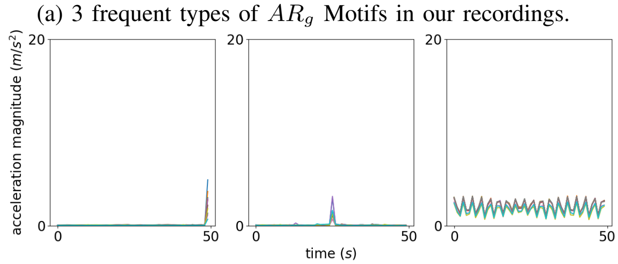
again employ the kNN classifier and the LOOCV technique for evaluation, as in the previous section.

Classification Results. We present the classification model's accuracy in the right segment of Table I. When considering each motif group individually, our model achieved a minimum accuracy of 53% and a recall of at least 67% ($m = 10$), reaching 64% and 74% for a pattern size of $m = 50$. Our scores significantly improved reaching 74% accuracy and, notably, 89% recall when aggregating Motifs from all visits for each infant. Similar to the previous task, correctly identifying AR infants with poor developmental outcomes remains a top priority to ensure they receive vital early care. Considering that infants are typically diagnosed with developmental disabilities around 24 months of age, our preliminary findings suggest that gathering longitudinal data across multiple sessions can contribute to the detection of developmental disabilities at much earlier ages.

Motif Interpretability. Figure 2 depicts the three most frequent Motif types identified in both classes. We observe that infants classified as AR$_g$ exhibit more complex patterns characterized by larger accelerations than AR$_p$ infants. The latter often show simpler patterns, often represented by single movements. Interestingly, the last motif in Fig.2b resembles the acceleration magnitude profile of a mechanical swing [13]. It's worth noting that our pipeline exclusively extracted such patterns from AR$_p$ recordings. We interpret their occurrence as a potential indicator of prolonged periods of inactivity, and thus, we have considered their presence in our analysis. To



(a) 3 frequent types of AR_g Motifs in our recordings.



(b) 3 frequent types of AR_p Motifs in our recordings. The last motif was described as the movement of a mechanical swing in [13].

Fig. 2: Plots of (a) AR_g and (b) AR_p Motifs in our recordings.

TABLE II: Average min, peak, mean acceleration, and repetition of extracted Motifs for *(i)* TD/AR and *(ii)* AR$_g$/AR$_p$ classes.

class	min acc.	peak acc.	mean acc.	repetition
TD	0.188	6.542	0.752	6.496
AR	0.137	6.171	0.357	11.338
AR_g	0.279	6.498	0.479	12.13
AR_p	0.047	6.022	0.215	13.15

further examine the Motif differences, we, again, computed the average *minimum*, *mean*, and *peak* of Motifs' acceleration, as well as their average *repetition* in the recordings for $m = 50$. The results are outlined in Table II. Motifs of AR$_g$ infants exhibit higher average acceleration min (p=0.019) and mean (p=0.023) compared to their AR$_p$ peers. The same trend is observed for the peak acceleration feature, although the difference is statistically insignificant (p=0.089). Additionally, the average pattern repetition is lower in AR$_g$ than in AR$_p$ infants, but this difference is also statistically insignificant (p=0.67). To the best of our knowledge, this study represents an initial attempt to compare specific patterns of at-risk infants for predicting their developmental outcomes, and therefore, these findings lack direct comparability with prior research.

IV. CONCLUSION & FUTURE WORK

This work advances prior research in infant movement monitoring by employing an algorithmic approach based on Motif extraction. Our analysis required no training to yield significant differences across the Motifs of infants from different developmental groups. Typically developing infants exhibited more complex and variable movement patterns than at-risk peers, and within the at-risk group, infants with poor developmental outcomes revealed simpler, singular movement patterns, in contrast to those with positive outcomes. These findings showcase the efficacy of Motifs as indicators of developmental disabilities earlier in infancy, even within a limited participant pool, presenting an alternative to current diagnostic tools that often identify such cases late. Future work involves expanding the dataset size to include more at-risk infants to comprehensively understand developmental assessment implications.

V. ACKNOWLEDGMENTS

This research has been funded in part by the NIH award R01LM014026. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors such as the NIH. This work was supported in part by the American Physical Therapy Association Section on Pediatrics Research Grant 1 and 2 Awards (to BAS). Smith's salary was supported in part, by the Foundation for Physical Therapy (NIFTI to BAS) and then in part by NIH (K12-HD055929 to KO).

REFERENCES

[1] A. S. Martin, R. W. Boles, L. Nocera, A. Kolatkar, M. May, Z. Hasnain, N. T. Ueno, S. Yennu, A. Alexander, and A. Mejia, "Objective metrics of patient activity: Use of wearable trackers and patient reported outcomes," 2018.

[2] J. M. Pevnick, K. Birkeland, R. Zimmer, Y. Elad, and I. Kedan, "Wearable technology for cardiology: update and framework," *Trends Cardiovasc. Med.*, vol. 28, no. 2, pp. 144–150, 2018.

[3] M. Lang, "Beyond fitbit: optical heart rate monitoring wearables and apps," *Alb. LJ Sci. & Tech.*, vol. 28, p. 39, 2017.

[4] M. T. Bianchi, "Sleep devices: wearables and nearables, informational and interventional," *Metabolism*, vol. 84, pp. 99–108, 2018.

[5] A. Hajisafi, M. D. Siampou, J. Bi, L. Nocera, and C. Shahabi, "Wearables for health (w4h) toolkit for acquisition, storage, analysis and visualization of data from various wearable devices," in *ICDE*, 2024.

[6] A. Hajisafi, H. Lin, Y.-Y. Chiang, and C. Shahabi, "Dynamic gnns for precise seizure detection and classification from eeg data," in *PAKDD*, 2024.

[7] S. E. Groen, A. C. E. De Blécourt, K. Postema, and M. Hadders-Algra, "General movements in early infancy predict neuromotor development at 9 to 12 years," *Dev. Med. Child Neurol.*, vol. 47, no. 11, pp. 731–738, 2005.

[8] M. Hadders-Algra and A. M. C. Groothuis, "Quality of general movements in infancy is related to neurological dysfunction, adhd, and aggressive behaviour," *Dev. Med. Child Neurol.*, vol. 41, no. 6, pp. 381–391, 1999.

[9] H. F. R. Prechtl, "State of the art of a new functional assessment of the young nervous system," *Early Hum. Dev.*, vol. 50, no. 1, pp. 1–11, 1997.

[10] M. C. Piper and J. Darrah, "Motor assessment of the developing infant," 1994.

[11] M. S. Abrishami, L. Nocera, M. Mert, I. A. Trujillo-Priego, S. Purushotham, C. Shahabi, and B. A. Smith, "Identification of developmental delay in infants using wearable sensors," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–7, 2019.

[12] N. T. Fitter, R. Funke, J. C. Pulido, M. J. Matarić, and B. A. Smith, "Toward predicting infant developmental outcomes from inertial motion recordings," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2305–2314, 2020.

[13] 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," *Sensors*, vol. 15, no. 8, pp. 19 006–19 020, 2015.

[14] I. A. Trujillo-Priego and B. A. Smith, "Kinematic characteristics of infant leg movements produced across a full day," *J. Rehabil. Assist. Technol. Eng.*, vol. 4, p. 2055668317717461, 2017.

[15] J. Lonardi and P. Patel, "Finding motifs in time series," in *Proc. 2nd Workshop Temporal Data Mining*, 2002, pp. 53–68.

[16] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh, "Matrix profile i: all pairs similarity joins for time series," in *2016 IEEE 16th ICDM*. IEEE, 2016, pp. 1317–1322.

[17] S. M. Law, "Stumpy: A powerful and scalable python library for time series data mining," *J. Open Source Software*, vol. 4, no. 39, p. 1504, 2019.

[18] B. A. Smith, D. L. Vanderbilt, B. Applequist, and A. Kyvelidou, "Sample entropy identifies differences in spontaneous leg movement behavior between infants," *Technologies*, vol. 5, no. 3, p. 55, 2017.