InfantMotion2Vec: Unlabeled Data-Driven Infant Pose Estimation Using a Single Chest IMU

Mohammad Nur Hossain Khan[1], Nancy L. McElwain[3, 4], Mark Hasegawa-Johnson[2, 4], and Bashima Islam[1]

[1]Department of Electrical and Computer Engineering, Worcester Polytechnic Institute
[2]Department of Electrical and Computer Engineering, University of Illinois
[3]Department of Human Development and Family Studies, University of Illinois
[4]Beckman Institute for Advanced Science and Technology, University of Illinois
mkhan@wpi.edu, mcelwn@illinois.edu, jhasegaw@illinois.edu, bislam@wpi.edu

2024 IEEE 20th International Conference on Body Sensor Networks (BSN) | 979-8-3315-3014-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/BSN63547.2024.10780750

***Abstract*—Early identification of neuro-developmental risks in infants is crucial for timely intervention and improved quality of life. Current screening methods are costly, intrusive, and limited by artificial environments or require the infant to wear multiple sensors. To address these challenges, we propose a novel approach leveraging inertial measurement units (IMUs) to monitor infants' spontaneous motor abilities in natural settings. Our method introduces a hierarchical semi-supervised classifier and the InfantMotion2Vec embedding to capture detailed motion patterns, accommodating a wide age range (up to 36 months) while minimizing reliance on labeled data and cumbersome sensor setups. We collected labeled IMU data from 25 families and unlabeled data from 42 families using a single wearable sensor. Pretraining an embedding network using unlabeled data with a hierarchical pose estimator resulted in a 26% increase in F1-score and a 77.7% increase in Cohen's Kappa score compared to using only labeled data. The InfantMotion2Vec embedding adequately handles highly unbalanced labeled data, demonstrating its effectiveness in infant posture classification.**

***Index Terms*—Self-supervised Model, Representation Learning, Hierarchical Classification, Posture Detection**

I. Introduction

Studies indicate that one in ten infants are at neuro-developmental risk due to factors such as premature birth, birth asphyxia, stroke, metabolic derangement, and intrauterine substance exposures [1]. While early therapeutic intervention can significantly improve their quality of life [1], identifying at-risk infants currently requires extensive and costly screening tests, including brain imaging [2]. These challenges highlight the need for scalable, effective, and robust early developmental screening techniques.

Infants' spontaneous movements offer vital insights into brain function [3]. The development of cognitive and language abilities in infants is intricately linked to early bodily movements, control of posture, and voluntary actions. By characterizing these spontaneous patterns over time, we can potentially identify neuro-developmental risks at an early stage. However, current assessments are often conducted in artificial or restricted environments like labs or hospitals, making it difficult to evaluate true neuro-developmental performance accurately and with ecological validity [4]. This further emphasizes the need for robust screening techniques that can be applied in natural, everyday settings. To address this need, recent studies have used inertial measurement units (IMUs) [5], [6] to monitor infants' spontaneous motor abilities in out-of-hospital settings over extended periods. These approaches aim to capture natural infant motor behavior more accurately, providing a more reliable assessment of neuro-developmental health.

However, prior studies using wearable sensors to assess infant postures and movements have been limited by several factors. First, the setup often involves multi-sensor wearables, which can be uncomfortable for the infant, making day-long assessments impractical. Second, these studies require large labeled datasets, which are difficult to obtain in home settings due to privacy concerns and the labor-intensive nature of data annotation. For instance, Airaksinen et al. [5] used multi-sensor wearables on infants' hands and legs to collect data from 24 infants (mean age 6.7 months, range 4.5–7.7) and identify five postures. However, their approach fails to account for task performance variations across ages, leading to imbalanced datasets. Similar studies involving a broader age group (4.5 to 19.5 months) [6] still face issues with cumbersome setups and large labeled dataset requirements. These shortcomings underscore the need for a new algorithm that provides accurate assessments with minimal intrusion and reduces reliance on extensive labeled data.

Human activity recognition has been an extensive research area for the last decade. Researchers have used hand-crafted features from IMU data as input to recurrent neural network models (RNN) to classify six adult activities, including sitting, standing, walking, or running [7], [8]. Some studies have also used semi-supervised learning and adversarial learning to learn the representation and then classify adult activities [9], [10]. However, adult activities are fundamentally different from child activities. Unlike adults, infants exhibit uncoordinated and unpredictable motion sequences, further complicating the data collection process. Thus, models trained on adult activity recognition fail to satisfactorily classifying child activities. Moreover, because simple approaches like the Madgwick orientation filter perform poorly in classifying postures by introducing addition jitter to our noisy data [11], and because infants' data are noisier compared with adults, a more complex

This work was supported by funding from NIDA (R34DA050256)

approach/model is needed.

Additionally, although use of a single chest-worn IMU sensor addresses a key limitation of prior work involving multiple-sensors setups, it poses several unique challenges for algorithm development. The variability in infant movement patterns, coupled with their rapid developmental changes, makes capturing consistent and accurate motion data difficult. Moreover, the single-sensor approach limits the spatial resolution and coverage of motion data, as it relies on a solitary point of reference. Nonetheless, the single-sensor approach offers the advantages of reduced hardware complexity and greater comfort for the infant, making it a compelling area of research.

This study introduces the first hierarchical semi-supervised classifier designed to identify three infant postures using a single chest-worn IMU during daylong home recordings. Utilizing data from children aged one to thirty-six months, it addresses movement variability by creating the novel InfantMotion2Vec embedding from unlabeled data. This embedding supports a pose estimation model using minimal labeled data, significantly reducing the need for multiple sensors and extensive datasets. Our approach offers a practical, scalable, and efficient solution for early neuro-developmental screening across diverse age groups.

For data collection, we employ a multi-modal wearable, LittleBeats™, placed in the chest pocket of a specially designed infant shirt [12]. This platform has previously been used for automatic speaker diarization and vocalization classification and environmental noise tagging in home [13], [14]. In addition to collecting audio and cardiac data, the LittleBeats device includes a single IMU sensor. For this report – the first to report on LittleBeats IMU data – we collected and labeled 1.07 hours of IMU data from 26 families with children of various ages and 1100 hours of unlabeled data from 42 families.

Our approach introduces a novel application of semi-supervised learning in the context of infant posture classification using a single IMU. Compared to other semi-supervised architectures typically used for adult human activity recognition, our framework performs better in classifying three distinct infant positions. This highlights our method's efficacy and innovation, specifically tailored to address the unique challenges of infant neuro-developmental assessment.

II. Data

A. Data Collection and Annotation

Families were recruited through various channels, including online listservs, flyers distributed to local organizations (e.g., libraries), and pediatric clinics. Parents received a complete description of the study procedures and provide written consent for their own and their child's participation. All study procedures are approved by the Institutional Review Board at the University of Illinois Urbana-Champaign.

B. Data Distribution

Unlabeled IMU Data. We collected 183 daylong recordings from 42 children (Mean age = 17.36 months, Range: 5.16 to 35.1 months), amounting to 1100 hours of data, which we use to develop a robust motion embedding, InfantMotion2Vec.

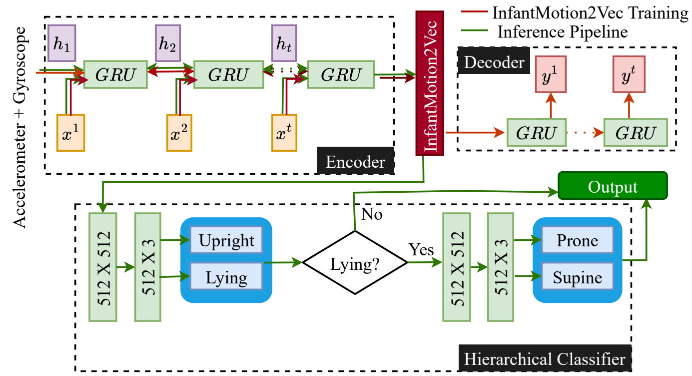


Fig. 1: Model architecture for InfantMotion2Vec and hierarchical pose estimation

Labeled Data. We collected IMU data from 25 children (Mean age = 19.95 months, Range: 1.1 to 47.5 months), with each child contributing approximately 30 minutes of labeled data. These children are different from those who contributed the unlabeled data. Mother-child dyads participated in a virtual visit via Zoom, which was video-recorded for labeling. Trained research assistants labeled the onset and offset times of behaviors in four categories (child posture/position, upper-body movement, lower-body movement, child-parent contact) using ELAN software [15]. This paper utilizes three infant positions: sitting or standing ("upright"), supine ("back"), and prone ("stomach"). We annotate 64 minutes of data (3186 seconds of upright, 507 seconds of back, and 150 seconds of stomach).

C. Validation Scheme

We employ an external validation scheme (leave-family-out) to evaluate our model's performance. We reserve data from five families for testing and use data from the remaining twenty families to train the classifier. This approach ensures that our model generalizes well across different subjects.

III. Infant Position Estimation with Semi-Supervised Learning

A. InfantMotion2Vec

InfantMotion2Vec is an infant motion embedding that learns the representation of infant posture and movement from unlabeled IMU data. Inspired by the Activity2vec framework [16], InfantMotion2Vec consists of a sequence-to-sequence network that maps the sequence of sensor data to itself, learning the features of the raw sensor data. Like Activity2Vec, InfantMotion2Vec has an encoder to learn the features by projecting the input into a fixed-dimension vector and a decoder to reconstruct the input from this vector. However, instead of a long short-term memory (LSTM) layer in Activity2Vec, InfantMotion2Vec uses a single-layer bidirectional Gated recurrent unit (GRU) for the encoder, and a unidirectional GRU for the decoder, as GRUs are more computational efficient with less parameter compared to LSTMs [17]. The network is trained by minimizing the mean square error (MSE) between

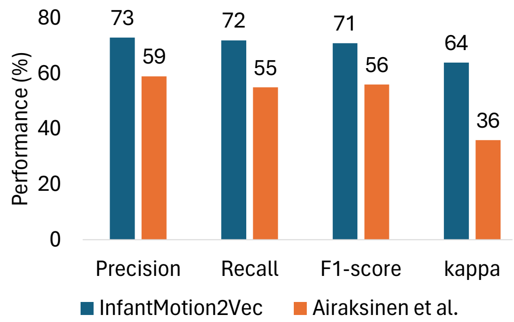


Fig. 2: InfantMotion2Vec and a hierarchical classifier outperform Airaksinen et al.'s supervised pose estimation [5].

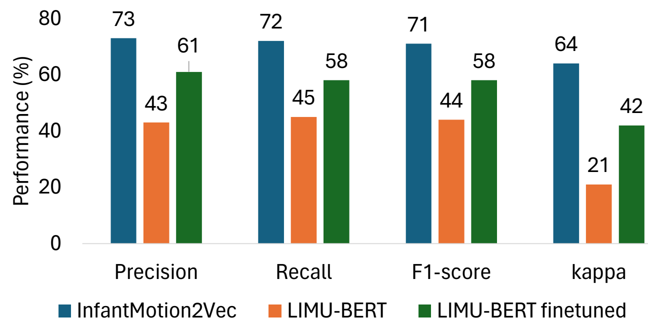


Fig. 3: InfantMotion2Vec outperforms both LIMU-BERT [9] and fine-tuned LIMU-BERT. All cases use the hierarchical classifier.

the original input and the decoder output. The framework architecture is shown in Figure1.

B. Hierarchical Pose Estimation

We develop a two-step hierarchical classifier (Fig 1) that takes the IMU data representation from InfantMotion2Vec and identifies three infant poses. For the hierarchical classification, we first consider lying on the back and lying on the stomach as the 'lying' class. Thus, we have a two-class classification problem where we differentiate between whether a baby is lying or upright. The output of the 'lying' class is then passed through another classifier of similar architecture to identify whether the baby is lying on its back or stomach. Each of these classifiers has two fully connected layers with 512 hidden units to classify an infant's position from labeled data. As InfantMotion2Vec already learns the data representation, we do not need a complex classifier for the final predictions, which could lead to overfitting and poor performance on unseen data.

C. Training Parameters

Using the Adam optimizer, we train the InfantMotion2Vec network for 1000 epochs. To reduce overfitting, we use an adaptive learning rate with a scheduler and L1-norm penalty. We use binary cross-entropy loss with the Adam optimizer and train for 50 epochs, saving the best model for classification. The training is performed on an NVIDIA RTX 3090Ti.

IV. Results & Discussion

We assess the performance of our pipeline, including InfantMotion2Vec and the hierarchical pose estimator, using precision, recall, F1 score, and Cohen's kappa score, followed by an ablation study of each component.

A. Classification Performance of Infant-Motion2Vec and Hierarchical Pose Estimator

We first evaluate the performance of the InfantMotion2Vec and hierarchical classifier, as shown in Figure 2. Our model achieves an F1-score of 71% and Cohen's kappa of 0.64. We also compare our framework's performance against state-of-the-art posture detection algorithms that use only labeled data [5]. Figure 2 shows that our model outperforms the model by Airaksinen et al., with a 26% and 77.7% increase in F1-score and Cohen's kappa, respectively. This indicates that a limited amount of highly unbalanced labeled data from a single sensor is insufficient to achieve satisfactory performance, even with a highly parameterized and complex network.

B. Performance of Embedding Network: Infant-Motion2Vec

We evaluate the performance of our embedding network, InfantMotion2Vec, by comparing it against the LIMU-BERT architecture [9]. To maintain consistency, we use a hierarchical classifier with LIMU-BERT. LIMU-BERT leverages a BERT-like attention model [18] to learn the representation of IMU sensor data and is trained on 80% of data from four different human activity recognition datasets: HHAR [19], UCI-HAR [8], Shoaib [20], and MotionSense [21]. Because LIMU-BERT lacks infant data, we fine-tuned it with labeled LittleBeats training data. Figure 3 shows that our model outperforms LIMU-BERT without fine-tuning, achieving a 69.7% and 20.4% higher F1 score and Cohen's kappa score, respectively. We also find that fine-tuned LIMU-BERT performs better than the original LIMU-BERT. However, InfantMotion2Vec achieves better performance than fine-tuned LIMU-BERT, with a 19.6% higher F1 score and 52.4% higher Cohen's kappa score. This indicates that embeddings learned from adult activities cannot accurately represent infant motion and position, and their performance degrades with highly unbalanced child data. Additionally, Figures 2 and 3 show that pre-trained LIMU-BERT outperforms the model used by Airaksinen et al., highlighting the importance of learning the representation of infant posture to improve classification performance when labeled data is limited.

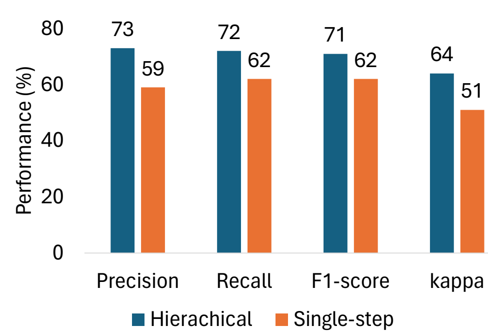
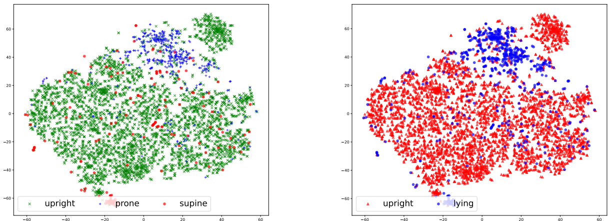


Fig. 4: With InfantMotion2Vec embedding hierarchical classifier outperforms the single-step classifier.



(a) upright vs. prone vs. supine (b) upright vs. lying

Fig. 5: t-SNE plot of InfantMotion2Vec embeddings

C. Performance of Hierarchical Pose Estimator

For this comparison, we use only the InfantMotion2Vec embedding and compare the performance of a single-step, three-class classifier with a hierarchical approach. Figure 4 shows that the hierarchical approach improves the F1 score by 14.5% and the Cohen's kappa score by 25.4%. Both the hierarchical and single-step classifiers outperform the model by Airaksinen et al., which uses only labeled data and shows the effectiveness of InfantMotion2Vec embedding. Figure 5 shows the t-SNE plot of InfantMotion2vec embeddings for all three classes and two classes after merging supine and prone classes into a single class, which shows that it is easier to detect two classes instead of three classes in a single step.

V. Conclusion & Future work

In this study, we use a single chest-worn IMU sensor for infant pose estimation, simplifying hardware setup and improving the comfort and practicality of monitoring infants compared to traditional multi-sensor methods. Our method also leverages unlabeled data to create an embedding, termed InfantMotion2Vec, which captures the complex and varied patterns of infants' positions with high fidelity without the need for large annotated datasets. Furthermore, integrating this embedding with a hierarchical classifier allows for satisfactory pose estimation, demonstrating that effective monitoring can be achieved with significantly reduced sensor data. Our approach outperforms previous multi-sensor infant pose estimation models and single-sensor adult activity recognition embeddings, highlighting the effectiveness of our method in capturing infant-specific motion patterns and positions.

In the future, we aim to develop a more sophisticated and robust model to improve classification performance and to identify more discrete positions and movements. Further, exploring additional sources of data and incorporating more advanced machine learning techniques could enhance the accuracy and reliability of our model. Valid assessment of infant motion may provide significant opportunities for the early detection and intervention of infants' motor and neurodevelopmental issues.

References

[1] S. A. Rosenberg, D. Zhang, and C. C. Robinson, "Prevalence of developmental delays and participation in early intervention services for young children," *Pediatrics*, vol. 121, no. 6, pp. e1503–e1509, 2008.

[2] I. Novak, C. Morgan, L. Adde, J. Blackman, R. N. Boyd, J. Brunstrom-Hernandez, G. Cioni, D. Damiano, J. Darrah, A.-C. Eliasson *et al.*, "Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment," *JAMA pediatrics*, 2017.

[3] C. Peyton, E. Yang, M. Msall, L. Adde, R. Støen, T. Fjørtoft, A. Bos, C. Einspieler, Y. Zhou, M. Schreiber *et al.*, "White matter injury and general movements in high-risk preterm infants," *American Journal of Neuroradiology*, vol. 38, no. 1, pp. 162–169, 2017.

[4] N. Nazneen, A. Rozga, C. J. Smith, R. Oberleitner, G. D. Abowd, R. I. Arriaga *et al.*, "A novel system for supporting autism diagnosis using home videos: Iterative development and evaluation of system design," *JMIR mHealth and uHealth*, vol. 3, no. 2, p. e4393, 2015.

[5] M. Airaksinen, O. Räsänen, E. Ilén, T. Häyrinen, A. Kivi, V. Marchi, A. Gallen, S. Blom, A. Varhe, N. Kaartinen *et al.*, "Automatic posture and movement tracking of infants with wearable movement sensors," *Scientific reports*, vol. 10, no. 1, p. 169, 2020.

[6] M. Airaksinen, A. Gallen, A. Kivi, P. Vijayakrishnan, T. Häyrinen, E. Ilén, O. Räsänen, L. M. Haataja, and S. Vanhatalo, "Intelligent wearable allows out-of-the-lab tracking of developing motor abilities in infants," *Communications Medicine*, vol. 2, no. 1, p. 69, 2022.

[7] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz *et al.*, "A public domain dataset for human activity recognition using smartphones." in *Esann*, vol. 3, 2013, p. 3.

[8] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.

[9] H. Xu, P. Zhou, R. Tan, M. Li, and G. Shen, "Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 220–233.

[10] M. Z. Zadeh, A. Ramesh Babu, A. Jaiswal, M. Kyrarini, and F. Makedon, "Self-supervised human activity recognition by augmenting generative adversarial networks," in *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*, 2021, pp. 171–176.

[11] J. Zevering, A. Bredenbeck, F. Arzberger, D. Borrmann, and A. Nüchter, "Imu-based pose-estimation for spherical robots with limited resources," in *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2021, pp. 1–8.

[12] N. L. McElwain, M. C. Fisher, C. Nebeker, J. M. Bodway, B. Islam, and M. Hasegawa-Johnson, "Evaluating users' experiences of a child multimodal wearable device: Mixed methods approach," *JMIR Human Factors*, vol. 11, p. e49316, 2024.

[13] J. Li, M. Hasegawa-Johnson, and N. L. McElwain, "Towards Robust Family-Infant Audio Analysis Based on Unsupervised Pretraining of Wav2vec 2.0 on Large-Scale Unlabeled Family Audio," in *Proc. INTERSPEECH 2023*, 2023, pp. 1035–1039.

[14] M. N. H. Khan, J. Li, N. L. McElwain, M. Hasegawa-Johnson, and B. Islam, "Sound tagging in infant-centric home soundscapes," in *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2024, pp. 142–146.

[15] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "Elan: A professional framework for multimodality research," in *5th international conference on language resources and evaluation (LREC 2006)*, 2006, pp. 1556–1559.

[16] A. Ghods and D. J. Cook, "Activity2vec: Learning adl embeddings from sensor data with a sequence-to-sequence model," *arXiv preprint arXiv:1907.05597*, 2019.

[17] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[19] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition," in *Proceedings of the 13th ACM conference on embedded networked sensor systems*, 2015, pp. 127–140.

[20] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10 146–10 176, 2014.

[21] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Mobile sensor data anonymization," in *Proceedings of the international conference on internet of things design and implementation*, 2019, pp. 49–58.