IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 30, 2019, accepted March 18, 2019, date of publication April 15, 2019, date of current version April 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911450

Discriminative Models of Spontaneous Kicking Movement Patterns for Term and Preterm Infants: A Pilot Study

KATELYN E. FRY[1], (Student Member, IEEE), YU-PING CHEN[2], AND AYANNA HOWARD[3], (Senior Member, IEEE)

[1]Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2]Department of Physical Therapy, Georgia State University, Atlanta, GA 30302, USA
[3]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA

Corresponding author: Katelyn E. Fry (katelyn.fry@gatech.edu)

This work was supported in part by the National Science Foundation (NSF) under Grant 1545287, and in part by the Linda J. and Mark C. Smith Endowed Chair.

ABSTRACT In this paper, we discuss machine learning methods for classifying gross kicking activity for the term and preterm infants. We examine different combinations of sensors to determine the relative importance of each sensor to gross activity detection. In addition, we discuss methods to correlate infant age to the amount of time an infant performs unilateral vs bilateral kicking and time an infant is at rest. For preterm infants, we examine this same relationship using birth age and adjusted age. From this comparison, we aim to determine which age is a better predictor for movement breakdown. For gross activity recognition, it was determined that a sensor placed on the thigh was less important to overall recognition than a sensor placed on the foot or shin. In addition, a sensor placed on the foot tended to be the most accurate on its own while the thigh sensor tended to be the least accurate. For the relationship between infant age and movement breakdown, it was determined that the amount of time spent at rest increases as age increases. Furthermore, the amount of time spent performing bilateral kicking decreases at a more rapid rate than unilateral kicking as age increases. Finally, we examine how this relationship changes over time for infants observed over multiple months.

INDEX TERMS Activity recognition, clustering methods, medical robotics, rehabilitation robotics, sensor fusion, wearable sensors.

I. INTRODUCTION

According to the Center for Disease Control and Prevention (CDC), one out of every ten births in the United States is considered premature (infants with a gestation period of 37 weeks or less). Despite the improved neonatal care, which has led to an increase in the number of surviving preterm infants, preterm infants are at an increased risk of developing neuro-developmental disorders. The most common of these motor disorders among children is cerebral palsy (CP) which affects approximately 2 to 2.5 per 1000 live births [1].

CP is a spectrum disorder that encompasses various categories of motor function disorders and varies in severity across individuals. Research has shown that interventions may improve the overall quality of life of affected individuals if CP can be reliably detected early in life [2], [3]. However, due to the variability between individual cases of CP, it is difficult to design a diagnostic test to encompass all patients [4]. Currently, detecting the development of CP in infancy requires clinical observation and documentation of functional motor milestones in combination with neurological assessments [5]. These approaches are not appropriate for the detection of CP early in life as such milestones are not typically exhibited in the first few months. Additionally, the observation of infant motor ability is typically confined to a clinical setting which limits the amount of time an infant can be observed [6]. Moreover, these approaches are subjective by nature due to their dependence on infant cooperation during the observation time and opinion of the clinician. To date, an objective method for the extended observation of

The associate editor coordinating the review of this manuscript and approving it for publication was Qingxue Zhang.

infant motor development and the early detection of CP is not available.

Spontaneous kicking is one of the earliest displays of motor skills and is an important precursor to later voluntary motor control [7]. Abnormal neuromotor function later in life is indicated by abnormal neuromotor function displayed through spontaneous kicking [8], [9]. The direct observation of an infant's spontaneous kicking early in life can be used to detect the development of neurodevelopmental disorders like cerebral palsy (CP). However, abnormalities in spontaneous movements are not well defined and are not readily observable through traditional functional motor milestone assessments.

FIGURE 1. Example of sensor placement for infant's left leg.

Our research aims to develop a system for the early detection of delays in infant motor development and developing CP through the extended observation of infant spontaneous kicking outside of the clinical setting. Infant kicking kinematic data is gathered over long periods of time using sensors attached to the limb segments of the infant's legs (Figure 1). This data is used to compute kicking kinematics and determine various features of infant spontaneous kicking. These features are used to determine a kinematic suite that describes the characteristics of typical spontaneous kicking at different months of development. With this suite, an infant is considered motor developmentally delayed if their spontaneous kicking displays characteristics typical of a younger infant.

In this work, we focus on detecting periods of rest, unilateral activity, and bilateral activity. We examine the ability of different machine learning algorithms to identify periods of kicking activity when different combinations of sensors are present. From this, we determine the most accurate classifier for this application and determine the ability of different combinations of sensors to determine kicking activity. Additionally, we relate infant age to the movement breakdown. A simple linear regression is fit to this relationship to develop a discriminative model to predict infant developmental age from movement breakdown.

II. RELATED WORK

A. AUTOMATED ANALYSIS OF INFANT MOTOR DEVELOPMENT

Numerous techniques have been developed to observe and analyze various aspects of infant motor development. Multiple approaches involve the use of specialized equipment like depth cameras and motion tracking systems to gather infant movement data. Depth cameras have been used to gather infant movement data for analyzing kinematic motion and for infant pose estimation [10]–[12]. Olsen *et al.* used motion capture to gather infant pose data and track infant movement [13]. Karch *et al.* used electromagnetic tracking to examine upper and lower limb motions of infants [14]. Electromagnetic tracking was also used in [15] to track fidgety movements in 3D space. The methods that utilize this specialized equipment provide precise spatial tracking of infant spontaneous movement. However, such approaches and others like them require a controlled environment and expensive equipment, making these approaches unsuitable for use outside of a laboratory or clinical setting.

Other approaches utilize optical devices like video cameras to analyze infant motion data. Oftentimes, markers are placed on the infant to aid in tracking the infant's movements. In [16], baseline data of early spontaneous movement in preterm infants were established from the kinematic analysis of video data with joint markers on the infants' lower limbs. Other approaches attempt to track infant motion and identify at risk infants using marker-less video data [17]. Adde*et al.*utilized motiongrams to capture general movement patterns without the need of markers [18]. Stahl *et al.*used motion trajectories from markerless video data to analyze spontaneous movements using optical flow and wavelet analysis [19]. Das *et al.*tracked infant kicking from video data to collect kinematic data to identify periods of simultaneous and non-simultaneous movements [20]. However, these methods often make assumptions about the position of the joints and require a specific configuration between the camera and the infant being filmed. Additionally, these methods are oftentimes not robust to occlusions making these approaches non-ideal for usage outside of a clinical or laboratory setting.

Other approaches use wearable technology or clothing embedded with sensors to gather infant movement data. Smith *et al.* determined the daily quantity of infant leg movement from infant kicking data gathered over the course of a full day to determine the daily kicking sequence [21]. Accelerometer data was used in [22] to identify motor milestones in infants and identify at risk infants. Due to their low cost and ability to be utilized in multiple settings, these approaches are suitable for in-home usage. However, these approaches are limited as they require a strict definition of which movements qualify as a kick.

To enable the early detection of delays in infant motor development and developing CP, a system must allow for a longer observation time than typically allowed in a clinical setting. In our methodology, we utilize the advantages of

wearable technology to enable observation in the home and thus maximize observation time. Additionally, we address the shortcomings oftentimes associated with wearable technology in this space. Rather than restricting our analysis to movements that follow a strict definition, our methodology allows for the analysis of a multitude of spontaneous movements. These movements could aid in the determination of developmental age and would otherwise be discarded.

B. ACTIVITY DETECTION FROM MULTIPLE SENSORS

There are multiple approaches available to combine information gathered from multiple sensors and derive an overall decision. The first subset of these approaches is called sensor fusion, in which features from the data of multiple sensors are combined into a feature vector that is used to reach an overall decision [23]. The second subset of these approaches is called decision fusion, in which the decisions of multiple sensors are combined to reach an overall decision. Traditional sensor fusion methods assume that the feature vectors are complete; that is, it is assumed that data is not missing. Though there exist methods to estimate missing data to form complete feature vectors, these methods can lead to large errors in overall classification [24]. In comparison, data fusion methods tend to be more robust to missing data. To account for a missing sensor decision, the rule to reach an overall decision can be easily adjusted. However, the optimal rule to combine individual sensor decisions is not always known.

Fry et al. used a Stance Hypothesis Optimal Detector (SHOD) with an automated thresholding method to determine instances of activity for term and low-risk preterm infants. Identified periods of activity for the individual sensors were then combined to determine overall activity for the leg [25]. In that study, two methods were considered to determine the overall activity for the leg. The first method, OR, dictated an instance of activity if a single sensor detected activity. This method tended to overestimate the instances of activity and the activity detector had a relatively large number of false positives which negatively impacted the overall classification accuracy. To address this issue, the second method, MODE, dictated an instance of activity if two or more sensors detected activity. The number of false positives decreased using the second method, however there was no significant difference between the accuracy of the two methods. From this work, it was determined that a better understanding of the importance of individual sensors to overall activity detection was needed. With this information, a smarter decision fusion/sensor fusion approach could be determined to significantly improve the accuracy of the detector.

To push forward the state-of-the-art in this domain, our current work discusses multiple machine learning assessment methods for identifying periods of kicking activity extracted from infant kicking data. Specifically, we aim to determine the relative importance of each sensor to gross activity detection for data acquired from term and low-risk preterm infants. Additionally, we discuss methods that correlate infant age to the amount of time an infant performs unilateral vs bilateral kicking. For preterm infants, we examine this same relationship using birth age and adjusted age. Finally, we examine how this relationship changes over time for infants observed over multiple months.

III. METHODS

To enable the early detection of neuro-developmental motor disorders in children, a system must:

- allow for a longer observation time than typically allowed in a clinical setting
- be adaptable to variations in infant size and age
- provide an objective, quantifiable metric of infant motor development.

The system proposed in this study uses an infant sensor suit to gather motion data associated with an infant's spontaneous kicking patterns. Collected data is then analyzed to determine instances of kicking activity as the first measure for calculating infant kicking kinematic data over long periods of time. The experimental procedures involving human subjects described in this section were approved by the Institutional Review Boards of the Georgia Institute of Technology and Georgia State University. Parents of the infants consented to the experimental procedures.

A. INFANT SENSOR SUIT AND DATA COLLECTION APP

Our system couples a Bluetooth-connected infant sensor suit with a data collection app, resident on a mobile device to enable ease-of collection in the home. The infant sensor suit pairs infant pants and rattle socks with six 6-axis IMU sensors powered by a coin cell battery (MbientLab's MetawearC). The suit incorporates 3 sensors per leg, placed on the thigh, shin, and foot to gather 3-axis acceleration and gyroscope data for each of the limb segments (Fig. 1) [37]. These sensors utilize Bluetooth Low Energy (BLE) technology for data transfer to our custom app. The app allows clinicians or parents to gather data from the sensor suit while monitoring battery life and connectivity of each sensor.

B. DATA COLLECTION PROCEDURE

Six infants, one male and five females, aged between 2 and 8 months (0.5 and 8 months adjusted age) were observed for this study (Table 1 and 2). Three of the eight infants were born full term while the remaining three infants were born premature but considered low risk. Low risk, preterm infants were defined as infants born at a gestational age between 32 to 37 weeks with no severe respiratory distress during birth and no existing Grade III or IV intraventricular hemorrhage after birth [24]. For observation in the home, infants were placed supine on a flat, padded surface while wearing the infant sensor suit. The infant's legs were momentarily held stationary at the beginning of each kicking session. For this baseline study, this step allowed us to later calibrate a zero-point time stamp with respect to quantifying the performance of our algorithms.

The infant was then encouraged to kick by providing stimulation consisting of verbal gestural cues and presentation of

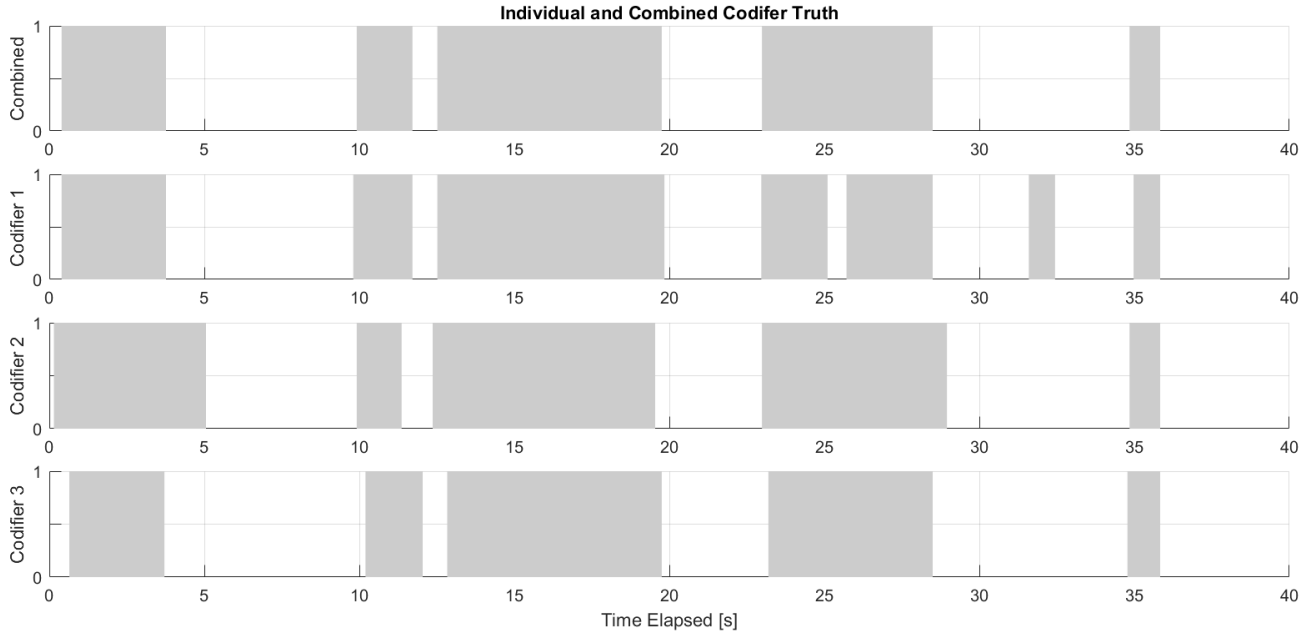


FIGURE 2. Example of individual codifier decision (bottom three subplots) and the overall combined truth vector. At instances of disagreement between the codifiers, the most popular identifier at a particular time was chosen to construct the combined truth vector. For example, at time elapsed t = 25 and t = 32, there was disagreement between codifier 1 and codifiers 2 and 3. As such, for the combined truth vector, the decision of codifier 2 and 3 were chosen over the decision of codifier 1.

TABLE 1. Infant demographics.

Infant	*Sex*	*Gestation Age (weeks)*	*Number of Sessions*
B	F	38	2
C_1	F	32	1
C_2	M	32	1
D	F	38	3
E	F	39	2
F	F	34	4

Gestation age is the number of weeks the infant was in the womb prior to birth. Infants with a gestation age less than 37 weeks are considered preterm. Number of sessions indicates the number of times an infant was sampled. Subscripts on infant identifier denote multiple births.

TABLE 2. Infant Session Information.

Infant	*Session Date*	*Session Length*	*Age at Testing (Months)*	*Adjusted Age (Months)*
B	09/14/2017	14m 56s	3	-
	10/26/2017	23m 30s	4.5	-
C_1	10/14/2017	14m 39s	5	0.5
C_2	10/14/2017	5m 11s	5	0.5
D	11/03/2017	10m 59s	2	-
	02/08/2018	17m 59s	5	-
	05/15/2018	13m 35s	8	-
E	12/12/2017	15m 25s	2	-
	01/30/2018	9m 18s	3.5	-
F	12/24/2017	5m 8s	2	0.5
	04/08/2018	16m 45s	5	3.5
	06/03/2018	13m 36s	7	5.5
	07/01/2018	8m 30s	8	6.5

Ages are rounded to the nearest half month. Adjusted age is only calculated for infants born preterm). Subscripts on infant identifier denote multiple births.

physical play objects. Stimulation was provided until it was determined that the infant needed to rest. That is, the kicking session continued until the parent or clinician requested a rest period or if the infant showed any form of agitation or distress. During kicking, acceleration and angular rate data were collected at a sampling rate of 100Hz from the embedded infant suit sensors. Several periods of data were collected with each session yielding up to 20 minutes of kicking data. Session length varied between infants due to infant emotional state; the average session length was 14 minutes. Each session was also filmed using a video timestamping application for the creation of truth data.

C. CONSTRUCTION OF GROUND TRUTH

Three independent coders were separately tasked to codify instances of leg motion activity in the timestamped videos. Each coder was asked to identify moments of time when the infant's leg was moving. Results from the coders were resampled to match the sampling rate of the data collection app and then used to construct a truth vector quantifying motion activity for each kicking session. If there was disagreement in motion activity between the coders, the most popular identifier (2 out of 3) at that timestamp was used to construct the truth vector (Fig. 2).

Coders were not given any specific instructions regarding which limb segment motion to prioritize. They were also not given instructions regarding interferences to or physical influencers of the infant's movements. These instances include occurrences where another individual, such as the parent, physically moved the infant or the infant's legs during an observation session.

D. CONSTRUCTION OF FEATURE VECTOR FOR ACTIVITY DECTECTION

To account for local dependencies between data points, a sliding window with a width of 250 ms and step of 100 ms was used for each segment of data. These values were determined through previous experiments to optimize the accuracy of kicking detection while compensating for sensor noise. A SHOD was used to create the feature vectors for the activity detector. SHOD is a magnitude-based method that uses acceleration and angular rate data to increase the precision and accuracy of an activity detector [17]. The SHOD value for each window was calculated and associated with the timestamp value at the center of the window (also called an instance of data). Then, these values were combined to create the figure of merit v_j for each sensor per leg. An explanation of the computation of v_j and definition of constants is given in Appendix. The figures of merit for each sensor or combinations therein were used to create a feature vector V for the activity detector. A feature vector was constructed for each of the infant's legs for each window of data.

IV. GROSS ACTIVITY DETECTION

In this section, we determine the relative importance of each sensor to the overall detection results. We compare cases where all 3 sensors per leg are present, where data from two sensors are present, and where data from one sensor is present. We examine the ability of three different machine learning algorithms to identify periods of kicking activity with different combinations of sensors to ensure that any trends observed are not dependent upon the classifier used.

A. APPROACHES FOR ACTIVITY DETECTION

In this work, three approaches were used to detect periods of activity: a thresholding method, a K-nearest neighbors (KNN) supervised learning method, and a Gaussian mixture model (GMM) unsupervised learning method. For the thresholding method and the KNN supervised learning method, a leave-one-out approach was used for training. In the leave-one-out approach, a segment of data is left out during training and the remaining segments of data are used to train the classification model. Detection results are then reported on the left-out segment of the data to evaluate the ability of the model to generalize to unseen data. For the GMM unsupervised learning method, a model was created for each segment of data without providing truth labels with respect to the detection output for that segment. Specific details for each method are detailed as follows.

1) THRESHOLDING

A thresholding method assumes that a data set can be optimally separated into two distinct classes or groups. In this method, we determine a threshold γ from a set of training data that separates the data into two distinct classes. As γ cannot perfectly separate the two classes, γ is chosen to maximize a desired classification metric (e.g. accuracy, specificity, etc.) when data in each separated class is compared to the baseline truth data [26]. In the leave-one-out approach, γ for each segment of data is determined. Then the averaged γ across the segments of data is used as the threshold for the left-out segment.

In this work, γ was specified as the point of maximum efficiency for each of the remaining segments of data as computed from the receiver operating characteristic (ROC) curve. Efficiency is a weighted average of sensitivity and specificity. Thus, the point of maximum efficiency represents the cutoff that maximizes both sensitivity and specificity. Accuracy was not used as the specified metric due to the tendency of the threshold being biased based on the larger frequency of samples from one group over another. For example, the threshold would be biased when using accuracy as the optimized metric if the infant was not moving for a significant portion of the data segment.

In activity detection, the two groups for classification are as follows: a negative instance of activity (or no motion) and a positive instance of activity (or movement). For a given instance of testing data x_i, a positive instance of activity was indicated if:

$$\|V(x_i)\| > \gamma \tag{1}$$

where $V(x_i)$ is the feature vector associated with instance x_i. If the above condition was not satisfied, a negative instance of activity was indicated for x_i. $\|V(x_i)\|$ was calculated by taking the Euclidean norm.

2) K-NEAREST NEIGHBORS (KNN)

The KNN method is an unsupervised, non-parametric method for classification. KNN uses the training dataset directly to make predictions for new unseen data. For a new instance x_i, a prediction is made by searching through the entire training set for the K most similar instances or neighbors [27]. A distance metric is used to determine which instances in the training dataset, y, are most similar to the new instance x_i. For this work, the distance metric, $d(x_i, y_r)$, is defined using a Euclidean distance:

$$d(x_i, y_r) = \sqrt{\sum_{j=1}^{n} \left(v_j(x_i) - v_j(y_r)\right)^2} \tag{2}$$

where n is the number of dimensions of the feature vector (the number of sensors present) and y_r represents a single instance in the training dataset to which the testing instance x_i is being compared. the K nearest neighbors for x_i are the K instances from the training set with the smallest $d(x_i, y_r)$.

The predicted class for x_i is determined as the class from the training set that had the highest frequency from the K most similar instances. That is, the class from the training set with the majority of the K nearest neighbors is taken as the prediction for the new instance. The classifier in this work uses the $K = 50$ nearest neighbors from the training set to determine the instance of activity for a given instance.

3) GAUSSIAN MIXTURE MODEL (GMM)

A GMM is a probabilistic model used to represent the presence of subpopulations, or groups, within a larger population. This method constitutes a form of unsupervised learning and thus does not require a set of labeled observed (training) data to identify these groups (also known as components of the GMM) [28], [29], [30]. GMM assumes that the individual feature vectors $V(x_i) \in V$ are derived from a mixture or sum of a finite number of Gaussian or normal distributions with unknown parameters. These individual distributions model the distribution of the data, the probability density functions (pdf's) within the different groups. The overall mixture model $p(V)$ has the form:

$$p(V) = \sum_{k=1}^{K} \phi_k N(V \mid \mu_k, \Sigma_k) \tag{3}$$

$$\sum_{k=1}^{K} \phi_k = 1 \tag{4}$$

where $N(V|\mu_k, \Sigma_k)$ is the pdf of group k, μ_k and Σ_k are the mean and covariance matrix specifying the normal distribution, and K is the number of groups. The group mixture weights, ϕ_k for group k, are constrained to sum to 1 so that the total pdf normalizes to 1. Finally, the dimension of μ_k and Σ_k are determined by the dimension of the feature vectors $V(x_i)$.

An expectation-maximization algorithm is used to iteratively estimate the model parameters $(\mu_k, \Sigma_k, \phi_k)$ for each normal distribution in the mixture model. The expectation step determines the expected group assignment C_k for each instance x_i is calculated given the model parameters (i.e. $p(C_k|V(x_i), \widehat{\mu}, \widehat{\Sigma}, \widehat{\phi})$). The maximization step then maximizes the expectations calculated in the expectation step and updates the model parameters. This process repeats until the algorithm converges resulting in a maximum likelihood estimate.

With the estimated model parameters, data can be clustered by assigning each datum to its most likely cluster assignment. That is, cluster assignment is by the most likely group assignment. The probability that an instance x_i belongs to a certain component assignment C_k is calculated by:

$$p(C_k \mid V(x_i)) = \frac{\phi_k N(V(x_i)|\widehat{\mu_k}, \widehat{\Sigma_k})}{\sum_{r=1}^{K} \phi_r N(V(x_i)|\widehat{\mu_r}, \widehat{\Sigma_r})} \tag{5}$$

where the model parameters specifying the normal distributions are the estimated parameters from the expectation-maximization algorithm.

In this work, a GMM is estimated such that the model clusters the data into two groups over N replicates or repetitions. Then, the estimated model most likely to describe the data is selected from the N replicates and used to assign instances of activity to each timestamp. $N = 15$ was chosen to ensure that at least one replicate converged while not being too computationally intensive.

B. ACTIVITY DETECTION WITH COMBINATIONS OF SENSORS

This work also aims to determine the robustness of the gross activity detection approaches relative to the presence or loss of sensor data acquired from term and low-risk preterm infants. As such, different combinations of sensor loss are considered when evaluating the three methods for activity detection to ensure that any trend observed is not dependent upon the classifier used.

1) ALL SENSORS PRESENT

We first examine the performance of the different activity detection approaches discussed in section IV. A. when all three sensors per leg are present. This serves as a baseline on the optimal performance associated with the three different approaches. In the baseline assessment, the individual sensor figures of merit are concatenated to create a three-dimensional feature vector at each instance (e.g. $V(x_i) = [v_{thigh}(x_i), v_{shin}(x_i), v_{foot}(x_i)]$).

2) TWO SENSORS PRESENT

In this section, we examine the impact a missing sensor has on overall activity recognition. That is, if data from one sensor is lost, how well does the algorithm perform with the remaining two sensors. We examine the case where the shin and thigh sensor are present, the case where the shin and foot sensor are present, and the case where the thigh and foot sensor are present. For the sensors present, the individual sensor figures of merit are concatenated to create a two-dimensional feature vector at each timestamp (e.g. $V(x_i) = [v_{thigh}(x_i), v_{shin}(x_i)]$).

3) ONE SENSOR PRESENT

We finally examine algorithm robustness by measuring its performance when only one sensor is present. This provides an indication of the reliability of the various approaches and their dependency on individual sensor placement. It also serves as an indication of the relative importance of each sensor to overall detection of motor activity.

C. RESULTS FOR GROSS ACTIVITY DETECTION

Table 3 depicts how well each algorithm predicts overall leg movement using the different combinations of available sensors. Values displayed are the average performance of the classifier over all infants over all testing sessions. Though no statistically significant claims can be made due to the small sample size, there were numerous observed trends.

TABLE 3. Classifier performance when different combinations of sensors present.

Left Leg:

Sensor Combination	Threshold			KNN			GMM		
	Acc.	Spec.	Sens.	Acc.	Spec.	Sens.	Acc.	Spec.	Sens.
All Sensors	0.820	0.759	0.856	0.826	0.804	0.836	0.787	0.746	0.806
Shin and Thigh Sensors	0.817	0.758	0.854	0.822	0.795	0.833	0.789	0.768	0.794
Foot and Thigh Sensors	0.817	0.752	0.858	0.825	0.796	0.838	0.783	0.744	0.800
Foot and Shin Sensors	0.820	0.757	0.856	0.824	0.797	0.838	0.781	0.742	0.811
Foot	0.817	0.754	0.854	0.823	0.801	0.839	0.764	0.751	0.766
Shin	0.820	0.756	0.857	0.822	0.795	0.834	0.779	0.770	0.768
Thigh	0.808	0.742	0.850	0.812	0.794	0.822	0.772	0.775	0.764

Right Leg:

Sensor Combination	Threshold			KNN			GMM		
	Acc.	Spec.	Sens.	Acc.	Spec.	Sens.	Acc.	Spec.	Sens.
All Sensors	0.833	0.792	0.844	0.837	0.807	0.855	0.790	0.775	0.798
Shin and Thigh Sensors	0.823	0.773	0.843	0.827	0.798	0.842	0.783	0.793	0.775
Foot and Thigh Sensors	0.831	0.783	0.848	0.835	0.804	0.855	0.782	0.779	0.784
Foot and Shin Sensors	0.834	0.784	0.857	0.834	0.802	0.853	0.785	0.769	0.796
Foot	0.830	0.778	0.854	0.833	0.798	0.856	0.772	0.781	0.760
Shin	0.822	0.773	0.844	0.825	0.792	0.843	0.762	0.803	0.752
Thigh	0.816	0.766	0.834	0.817	0.789	0.831	0.754	0.803	0.728

Overall, the highest accuracy was achieved when all three sensors were present. For the threshold method, a comparable detection accuracy was reached when the foot and shin sensors were present. In general, the omission of the foot or shin sensor impacted accuracy more than the omission of the thigh sensor. Additionally, the foot sensor tended to be the most accurate sensor for detecting activity on its own though in some instances, the shin sensor was more accurate. In general, the thigh sensor was the least accurate for detecting accuracy on its own. In terms of gross activity recognition, the thigh sensor was found to be less important to overall recognition than the foot or shin sensor.

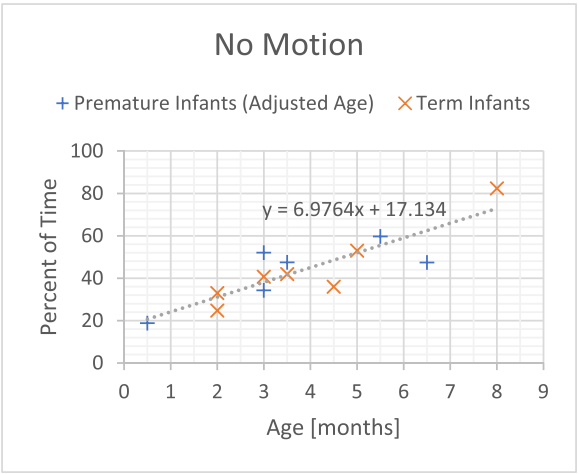
All methods and combinations of sensors have higher sensitivity than specificity. As such, all combinations of sensors were able to identify positive instances of activity better than negative instances of activity. Finally, of the three detection methods, the KNN method performed the best for all sensor combinations.

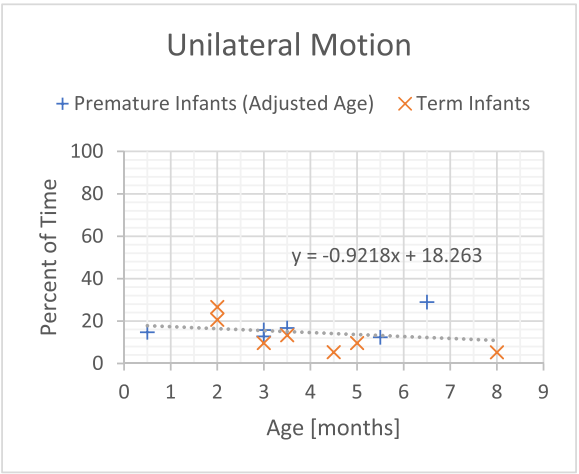
V. RELATIONSHIP BETWEEN MOVEMENT BREAKDOWN AND AGE

In this section, we discuss our method for determining correlations between characterization of the kicking activity and the infant's age. The objective for developing such a model is important in identifying features of normative kicking, leading to early identifications of delayed or atypical kicking profiles. For characterizing kicking activity, we decompose movements into four states and determine the proportion of time an infant spends in each of the states: at rest (no motion), unilateral motion (dominant leg movement only), and bilateral (both legs moving). In the unilateral motion state, the dominant leg depends on the individual. That is, percentages reported are either left leg unilateral motion or right leg unilateral motion depending on which leg is dominant for the infant (i.e. which state between unilateral left and unilateral right the infant spends more time in). Additionally, the development of preterm infants is generally measured on an adjusted scale. That is, preterm infants of a certain adjusted age are typically compared to term infants of that birth age. However, it is unclear whether a preterm infant should be considered by their adjusted age or their birth age in this relationship. In this study, we chose to evaluate preterm infants based on their adjusted age and their birth age.

A. DESCRIMINATIVE MODEL FITTING

To develop a model that defines the relationship between characterization of kicking activity and infant age, the percent of time the infant spent in each motion state was plotted





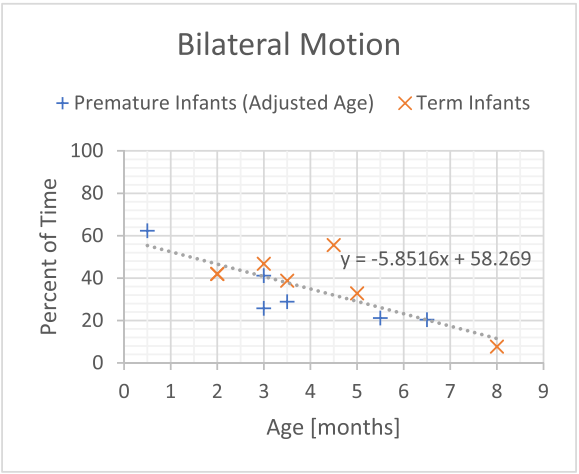
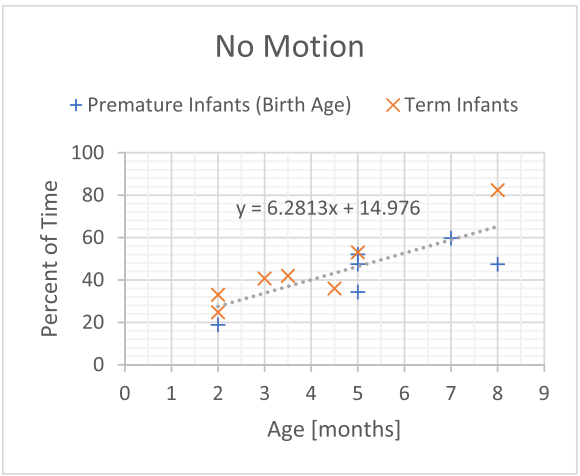
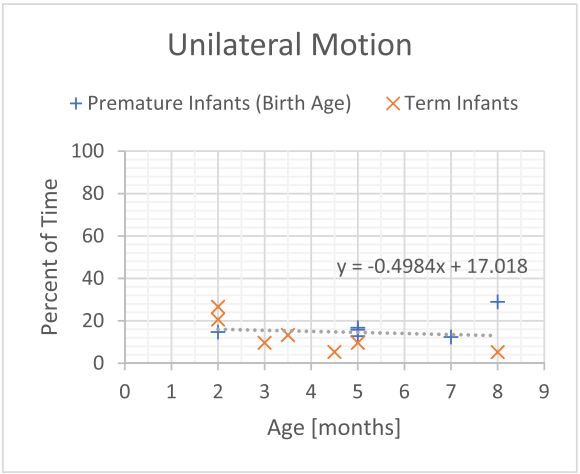


FIGURE 3. Plots for the percent of time an infant spent in the various motion states vs adjusted age rounded to the nearest half month. Here, the percentages for the premature infants (indicated with the blue + markers) were plotted against the infant's adjusted age. The adjusted age is calculated by subtracting the number of weeks the infant was born before their due date from the infant's birth age.





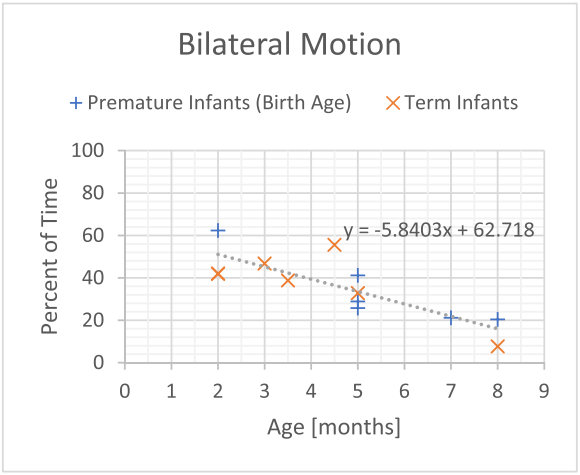
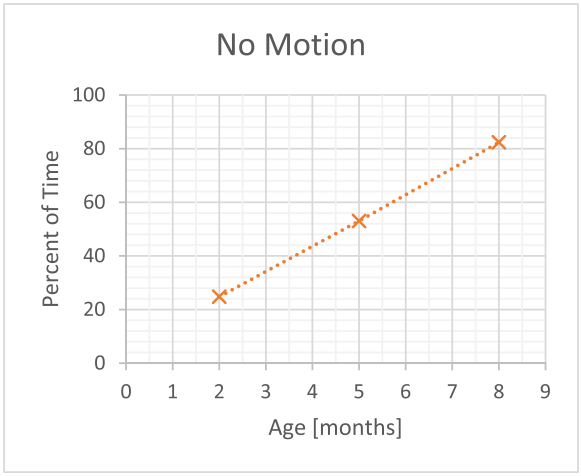
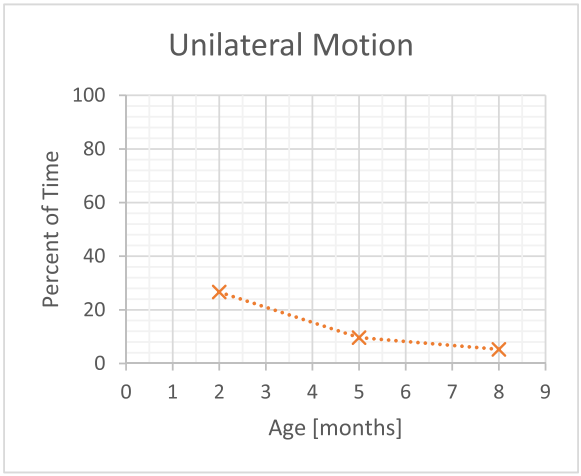


FIGURE 4. Plots for the percent of time an infant spent in the various motion states vs birth age rounded to the nearest half month. Here, the percentages for the premature infants (indicated with the blue + markers) were plotted against the infant's birth age. The motion states were fit with a linear model to determine which age premature infants should be considered when examining their movement breakdown.

against the infant's age. A linear regression was fit to each motion state to predict percentage of time given an infant's age. A sum of squares error (SSE) was then calculated for each motion state to indicate how well the linear model represented the data:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{6}$$





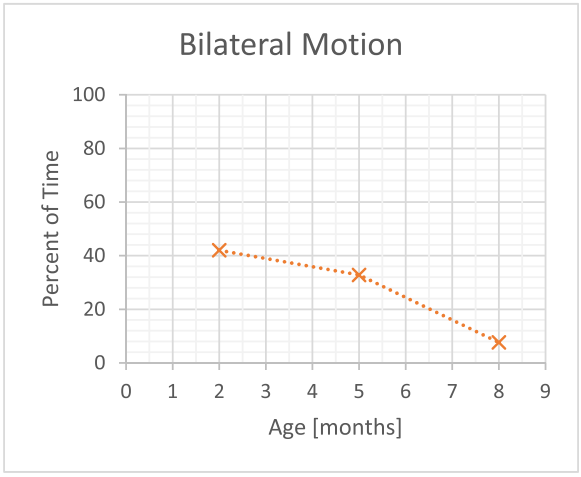
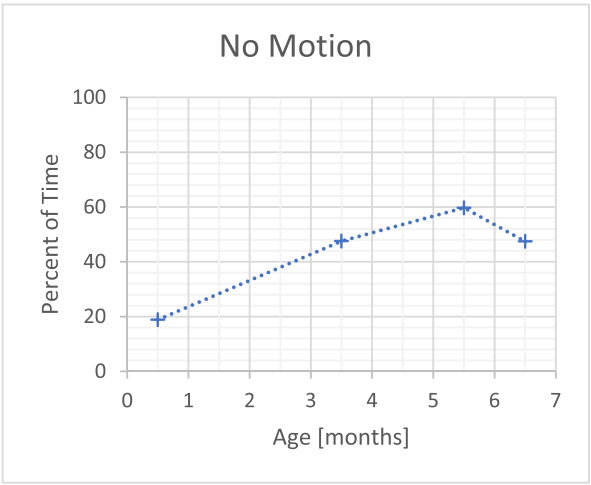
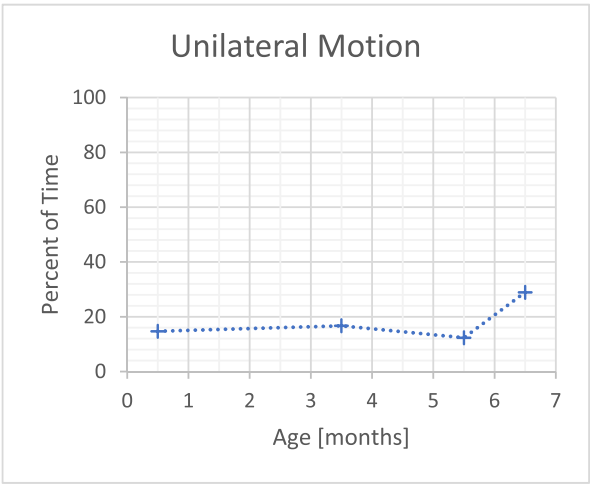


FIGURE 5. Plots for the percent of time infant D spent in the various motion states vs birth age. Infant D was born to term. For infant D, the general trends discussed for the entire group are observed here over multiple months. That is, the amount of time infant D spent at rest increases with age as the amount of time spent performing bilateral and unilateral motion decreased with age.





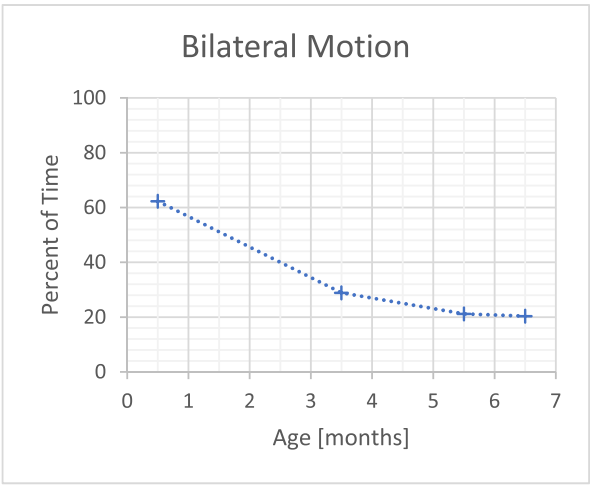


FIGURE 6. Plots for the percent of time infant F spent in the various motion states vs adjusted age. Infant F was born 6 weeks preterm. Infant F also follows the general trends discussed for the entire group. However, there is a distinct increase in % time for unilateral motion and a distinct decrease in % time for no motion from 5.5 month to 6.5 months adjusted age that goes against the trend for the overall group.

where y_i is an observed percentage and $\widehat{y_i}$ is the predicted percentage from the linear model. When the SSE's from two models are compared, the smaller SSE indicates the model with higher predictive power. The coefficient of determination, R^2, is also reported for each model. R^2 indicates the percentage of variability in the response variable that is explained

by the linear model. Generally, a higher R^2 indicates a better fitting model though this is not guaranteed. As such, it is important to consider R^2 in addition to another metric for goodness of fit like the SSE.

B. RESULTS FOR CORRELATING MOVEMENT BREAKDOWN AND INFANT AGE

Fig. 3 and 4 plot the observed percentage breakdowns vs infant adjusted age and infant birth age respectively. The breakdown of kicking activity reported is from activity as detected from the KNN classification method when all three sensors were present. Orange $\times$ markers represent observations from term infants while blue $+$ markers represent observations from preterm infants. Additionally, the linear models are displayed. In these results, we observe that the no motion class had a positive relationship with infant age while the unilateral motion class and the bilateral motion class had a negative relationship with age. Furthermore, the slope of the regression for the bilateral motion class is steeper than that of the unilateral motion class. Thus, as the infant ages, the amount of gross motor activity of the lower limbs decreases overall while instances of bilateral motion decrease more rapidly than instances of unilateral motion.

TABLE 4. Comparison of linear models for motion states.

Motion State		R^2	*SSE*
No Motion	Birth Age	0.66	1,071.93
	Adj. Age	0.75	791.60
Unilateral	Birth Age	0.02	703.26
	Adj. Age	0.07	651.445
Bilateral	Birth Age	0.67	904
	Adj. Age	0.62	1,047.04

Linear models displayed on their corresponding plots in Fig. 3 and 4. The predicted value for the linear model corresponds to the percent of time spent in a motion state while the predictor (age) corresponds to either the infant's birth age or adjusted age.

Table 4 reports the SSE's of the six linear models. Within each motion state, the SSE of the linear model associated with birth age is compared to that of the SEE of the linear model associated with adjusted age. Generally, the SSE for the models using the adjusted age of the premature infants was lower except for the bilateral motion state. Additionally, the R^2 value was higher for the adjusted age in the no motion state and the unilateral motion state while the R^2 value for the bilateral motion state was higher for birth age. From the current data, evaluating premature infants at their adjusted age generally results in smaller residuals and higher coefficients of determination and thus a more accurate prediction.

Fig. 5 plots the observed percentage breakdowns vs infant age for infant D and Fig. 6 plots the observed percentage breakdowns vs infant adjusted age for infant F. Infant D was born to term while infant F was born 6 weeks preterm. For infant D, the amount of time the infant spends at rest increases with age while the amount of time the infant spends performing bilateral or unilateral movement decreases with age. Similar trends are observed for infant F. However, there is a distinct increase in percent time for unilateral motion and a distinct decrease in percent time for no motion from 5.5 month to 6.5 months adjusted age that goes against the trend for the overall group.

VI. CONCLUSION

In this paper, we discussed machine learning methods for classifying gross kicking activity for term and preterm infants and examined different combinations of sensors to determine the relative importance of each sensor to gross activity detection. While the highest overall accuracy was achieved when the foot, shin, and thigh sensor were all present, our results indicate that for gross activity detection, the omission of a sensor placed on the thigh did not impact overall recognition as much as the omission of a sensor placed on the foot or shin. Individually, a sensor placed on the foot tended to be the most accurate on its own while the thigh sensor tended to be the least accurate.

We also discussed methods to correlate infant age to the amount of time an infant is at rest, performing unilateral activity, or performing bilateral activity. It was determined that as the amount of time they spent at rest increases as age increases. Furthermore, the amount of time spent performing bilateral kicking decreases at a more rapid rate than unilateral kicking as age increases.

Finally, we examined how this relationship changes over time for infants observed over multiple months. The trends observed for the three motion states when examining the group are reflected in the analysis of the two infants (one term and one premature) over multiple months.

To enable the early detection of delays in infant motor development and developing CP through extended observation of infant spontaneous kicking outside of the clinical setting. Characteristics of gross movement activity such as movement breakdown and the kinematic characteristics therein could serve as one feature to detect motor development delays.

APPENDIX

SHOD is a magnitude-based method that uses both acceleration and angular rate data to increase the precision and accuracy of an activity detector.

$$v(x_i) = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{\sigma_a^2}\left\|\boldsymbol{a}_k - g\frac{\bar{\boldsymbol{a}}_n}{\|\bar{\boldsymbol{a}}_n\|}\right\|^2 + \frac{1}{\sigma_\omega^2}\|\boldsymbol{\omega}_k\|^2\right) \quad (7)$$

where: n refers to a specific frame, centered at instance x_i, which represents the content (acceleration and angular rate data) of a sliding window; $\boldsymbol{a}_k$ and $\boldsymbol{\omega}_k$ are the acceleration and angular rate vector respectively for observation k of a specific frame n; $\bar{\boldsymbol{a}}_n$ is the mean of the acceleration vector of a specific frame n; g is the magnitude of acceleration due to gravity (1 g).

The constants determined are as follows: $N = 25$ is the number of samples in a frame determined by the desired length of the frame in seconds multiplied by the sampling frequency; $\sigma_a^2 = 3.24 \times 10^{-6}$, variance of the acceleration signal noise, and $\sigma_\omega^2 = 0.0049$, variance of the angular rate signal noise, are determined from specifications of the IMU.

ACKNOWLEDGMENT

Any opinions, findings, and conclusions or recommendations expressed in this research are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

[1] A. Herskind, G. Greisen, and J. B. Nielsen, "Early identification and intervention in cerebral palsy," *Developmental Med. Child Neurol.*, vol. 57, no. 1, pp. 29–36, 2015.
[2] E. Rogers *et al.*, "Smart and connected actuated mobile and sensing suit to encourage motion in developmentally delayed infants," *J. Med. Devices*, vol. 9, no. 3, Sep. 2015, Art. no. 030914.
[3] K. Subramanyam *et al.*, "Soft wearable orthotic device for assisting kicking motion in developmentally delayed infants," *J. Med. Devices*, vol. 9 no. 3, 2015, Art. no. 030913.
[4] D. Bryant *et al.*, "An infant smart-mobile system to encourage kicking movements in infants at-risk of cerebral palsy," in *Proc. IEEE Workshop Adv. Robot. Its Social Impacts (ARSO)*, Austin, TX, USA, Aug. 2017, pp. 1–6.
[5] M. Hadders-Algra, "Early diagnosis and early intervention in cerebral palsy," *Frontiers Neurol.*, vol. 5, p. 185, Sep. 2014.
[6] B. A. Smith *et al.*, "Daily quantity of infant leg movement: Wearable sensor algorithm and relationship to walking onset," *Sensors*, vol. 15, no. 8, pp. 19006–19020, 2015.
[7] J. P. Piek and N. Gasson, "Spontaneous kicking in fullterm and preterm infants: Are there leg asymmetries?" *Hum. Movement Sci.*, vol. 18, no. 2, pp. 377–395, 1999.
[8] F. B. Palmer, "Strategies for the early diagnosis of cerebral palsy," *J. Pediatrics*, vol. 145 no. 2, pp. S8–S11, 2004.
[9] M. Bax *et al.*, "Proposed definition and classification of cerebral palsy," *Developmental Med. Child Neurol.*, vol. 47, no. 8, pp. 571–576, 2005.
[10] S. Jeng *et al.*, "Kinematic analysis of kicking movements in preterm infants with very low birth weight and full-term infants," *Phys. Therapy*, vol. 82, no. 2, pp. 148–159, 2002.
[11] N. Hesse *et al.*, "Body pose estimation in depth images for infant motion analysis," in *Proc. Ann. Int. Conf.*, Aug. 2017, pp. 1–9.
[12] M. Serrano *et al.*, "Lower limb pose estimation for monitoring the kicking patterns of infants," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 2157–2160.
[13] M. D. Olsen, A. Herskind, J. B Nielsen, and R. R. Paulsen, "Using motion tracking to detect spontaneous movements in infants," in *Proc. Scandin. Conf. Image Anal.*. Cham, Switzerland: Springer, 2015, pp. 410–417.
[14] D. Karch *et al.*, "Kinematic assessment of stereotypy in spontaneous movements in infants," *Gait Posture*, vol. 36 no. 2, pp. 307–311, 2012.
[15] R. Parsa *et al.*, "Features for movement based prediction of cerebral palsy," M.S thesis, Dept. Eng. Cybern., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2009.
[16] C. B. Heriza, "Organization of leg movements in preterm infants," *Phys. Therapy*, vol. 68, no. 9, pp. 1340–1346, 1988.
[17] H. Rahmati *et al.*, "Video-based early cerebral palsy prediction using motion segmentation," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 3779–3783.
[18] L. Adde *et al.*, "Using computer-based video analysis in the study of fidgety movements," *Early Hum. Develop.*, vol. 85, pp. 541–547, May 2009.
[19] A. Stahl *et al.*, "An optical flow—Based method to predict infantile cerebral palsy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 605–612, Feb. 2012.
[20] D. Das, K. Fry, and A. M. Howard, "Vision-based detection of simultaneous kicking for identifying movement characteristics of infants at-risk for neuro-disorders," in *Proc. IEEE ICMLA*, Dec. 2018, pp. 1413–1418.
[21] B. A. Smith *et al.*, "Daily quantity of infant leg movement: Wearable sensor algorithm and relationship to walking onset," *Sensors*, vol. 15, no. 8, pp. 19006–19020, 2015.
[22] D. Gravem *et al.*, "Assessment of infant movement with a compact wireless accelerometer system," *J. Med. Devices*, vol. 6, no. 2, 2012, Art. no. 021013.
[23] A. K. Gao Lei Bourke and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems,"*Med. Eng. Phys.*, vol. 36, no. 6, pp. 779–785, 2014.
[24] P. J. Garcia-Laencina, J. Sancho-Gomez, and A. R. Figueiras-Vidal, "Pattern classification with missing values using multitask learning," in *Proc. Internation Joint Conf. Neural Netw.*, Jul. 2006, pp. 3594–3601.
[25] K. E. Fry, Y.-P. Chen, and A.Howard, "Detection of infant motor activity during spontaneous kicking movements for term and preterm infants using inertial sensors," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 5767–5770.
[26] A. Olivares *et al.*, "Detection of (in) activity periods in human body motion using inertial sensors: A comparative study," *Sensors*, vol. 12, no. 5, pp. 5791–5814, 2012.
[27] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.
[28] A. Moore. (2018). *Statistical Data Mining Tutorial on Gaussian Mixture Models*. [Online]. Available: http://www.cs.cmu.edu/ awm/tutorials
[29] B. Scherrer. (2007). *Gaussian Mixture Model Classifiers*. [Online]. Available: http://www. medialab.bme.hu/medialabAdmin/uploads
[30] D. Gravem *et al.*, "Assessment of infant movement with a compact wireless accelerometer system," *J. Med. Devices*, vol. 6, no. 2, 2012, Art. no. 021013.
[31] F. Attal, "Physical human activity recognition using wearable sensors," *Int. Fed. Autom. Control*, vol. 15, no. 12, pp. 31314–31338, 2015.
[32] B. Jeon and D. A. Landgrebe, "Decision fusion approach for multitemporal classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1227–1233, 1999.
[33] Seraji, Homayoun, and Navid Serrano, "A multisensor decision fusion system for terrain safety assessment," *IEEE Trans. Robot.*, vol. 25, no. 1, pp. 99–108, 2009.
[34] Chen, Biao, and Pramod K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Sampling Process.*, vol. 50, no. 8, pp. 1089–1818, Aug. 2002.
[35] Y. P. Chen and L. Fetters, "A comparison of the leg coordination patterns of preterm and fullterm infants: A meta-analysis," *Formosan J. Phys. Therapy*, vol. 27, no. 6, pp. 303–313, 2002.
[36] R. B. Lanjewar *et al.*, "Implementation and comparison of speech emotion recognition system using Gaussian mixture model (gmm) and k-nearest neighbor (k-nn) techniques," *Procedia Comput. Sci.*, vol. 49, pp. 50–57, Aug. 2015.
[37] *Detection of Infant Motor Activity During Spontaneous Kicking Movements for Term and Preterm Infants Using Inertial Sensor*, document 62700, Jul. 2018.

KATELYN E. FRY received the B.S. degree in electrical engineering from Clemson University, Clemson, SC, USA, in 2014, and the M.S. degree in robotics from the University of Michigan, Ann Arbor, MI, USA, in 2015. She is currently pursuing the Ph.D. degree in robotics with the Georgia Institute of Technology, Atlanta, GA, USA.

In 2016, she was a Researcher with the Savannah National Research Lab, Aiken, SC, USA. Her research area of interests includes healthcare robotics. In this domain, she is specifically interested in the development of rehabilitative and assistive robotics for children and of devices for use outside of the clinical setting (robots in the home).

Dr. Fry was a participant in the Accessibility, Rehabilitation, and Movement Science National Science Foundation Research Traineeship with Georgia Tech.

YU-PING CHEN was born in Taipei, Taiwan, in 1970. She received the B.S. degree in physical therapy from National Taiwan University, Taiwan, in 1992, and the Sc.D. degree in applied kinesiology from Boston University, Boston, MA, USA, in 2001.

From 2002 to 2005, she was a Lecturer with the Faculty of Physical Therapy, National Yang-Ming University, Taipei, Taiwan. She was a Postdoctoral Research Scientist with the Department of Psychology, University of Massachusetts, Amherst, MA, USA, from 2005 to 2007, and an Assistant Professor with the Department of Physical Therapy, California State University, Fresno, from 2007 to 2009. Since 2009, she has been with the Department of Physical Therapy, Georgia State University, Atlanta, GA, USA, as an Assistant Professor, and was later promoted to tenured Associate Professor, in 2015. Her research interests include examining the novel interventions, mainly virtual reality and robotics, to train children with disabilities, and early identifying infants with high-risks of motor disabilities.

Dr. Chen has served as a Manuscript Reviewer in various peer-reviewed journals, such as the *Physical Therapy, Disabilities and Rehabilitation*, the *PLOS One*, and the *Developmental Medicine and Child Neurology*.

AYANNA HOWARD received the B.S. degree in engineering from Brown University, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California. She is currently the Linda J. and Mark C. Smith Professor and the Chair of the School of Interactive Computing, Georgia Institute of Technology.

In 2013, she also founded Zyrobotics, an educational technology company focused on engaging children of diverse abilities. Prior to Georgia Tech, she was a Senior Robotics Researcher with the NASA's Jet Propulsion Laboratory. Her works focus on advancements in artificial intelligence (AI), assistive technologies, and robotics. To date, her unique accomplishments have been highlighted through a number of awards and articles, including highlights in *USA Today*, *Upscale*, and *TIME Magazine*, as well as being recognized as one of the 23 most powerful women engineers in the world by *Business Insider*.