IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Classifying Toe Walking Gait Patterns among Children Diagnosed with Idiopathic Toe Walking using Wearable Sensors and Machine Learning Algorithms

Rahul Soangra[1&2], Yuxin Wen[1], Marybeth Grant-Beuttler[2] and Hualin Yang[1]

[1]Fowler School of Engineering, Chapman University, Orange, CA 92866 USA

[2]Department of Physical Therapy, Crean College of Health and Behavioral Sciences, Chapman University, Irvine, CA 92618 USA

Corresponding author: Rahul Soangra (e-mail: soangra@chapman.edu).

This work was supported in part by the Children's Hospital of Orange County Foundation and Kay Family Foundation.

ABSTRACT Idiopathic toe walking (ITW) is a gait abnormality in which children toe touch at initial contact and demonstrate limited or no heel contact throughout the gait cycle. Toe walking results in poor balance, increased risk of falling, and developmental delays among children. Identifying toe walking steps during walking can facilitate targeted intervention among children diagnosed with ITW. With recent advances in wearable sensing, communication technologies, and machine learning, new avenues of managing toe walking behavior among children are feasible. In this study, we investigate the capabilities of Machine Learning (ML) algorithms in identifying initial foot contact (heel strike versus toe strike) utilizing wearable body sensors. Thirty-six children (Age 9.4±2.8 years) diagnosed with ITW participated in this study. Six ML algorithms, consisting of Support Vector Machines (SVM), decision tree (DT), random forest (RF), K-nearest neighbors (KNN), Multi-layer Perceptron (MLP), and Gaussian process (GP), could successfully classify initial contact walking patterns among ITW. We found that a simple KNN algorithm resulted in the highest accuracy of 92.92% and F1-score of 93.20% to differentiate toe walking gait versus best heel strike when using all four body sensors. We also found that toe walking resulted in higher variability in the sacral vertical accelerations among children diagnosed with ITW. Accurate quantification of toe walking steps in clinical applications is critical for assessing rehabilitation progress and designing new interventions for children diagnosed with ITW.

INDEX TERMS Best Heel Strike (BHS), Gait, Idiopathic Toe Walking (ITW), Inertial Measurement Unit (IMU), Machine Learning (ML)

I. INTRODUCTION

Toe walking is a gait abnormality described by clinicians as toe-to-toe touch initial contact of the foot [1]. Persistent toe-walking without treatment may lead to increased risk of tripping or falling [2], leg pain [3], impaired muscle and motor coordination [4], and develop structural abnormalities [5]. Children with autistic spectrum disorders, cerebral palsy, muscular dystrophy, intellectual disabilities, etc., have an increased likelihood of exhibit toe walking characteristics but toe walking also observed in healthy children with no signs of a neurological, orthopedic, or psychological condition [6, 7] and referred to as Idiopathic Toe Walking (ITW) [1]. At six years of age, approximately one out of every 20 children demonstrate ITW [8-10] worldwide.

Toe walking is usually identified visually by clinical experts. Furthermore, gait analysis utilizing laboratory-based motion capture systems provides an accurate, objective quantification of toe walking behavior [2, 11]. Standard laboratory-based gait analysis protocols require specialized laboratory equipment such as instrumented walkways, infra-red camera-based motion capture systems, or treadmills with embedded force plates. This laboratory assessment is expensive and limited since it requires specialized personnel to operate and analyze gait data. In addition, children with

ITW frequently modify their gait in a laboratory environment to a more normalized pattern. Thus, new systems which could detect toe walking in natural settings are needed. Currently, no commercial technology exists to identify toe walking (gait type) in real-world settings and intervene in real-time for gait correction. Rehabilitation of children diagnosed with ITW largely depends on clinicians' visual observations of the child's gait pattern or collecting feedback from parents for their gait [12], which are observational and rely heavily on the clinicians' experience and judgment [12] and are error-prone. We have earlier reported that pressure sensors [13] and machine learning algorithms [14] could identify toe strike events, and smart-insoles can objectively track toe strike gait [15] among children with ITW in their natural settings.

Moreover, even after the rehabilitative intervention, there is a tendency for these children to revert to toe walking gait. Thus, accurate information on the number of toe strikes and how long children walk on toes is critical. Hence, continuous gait pattern monitoring is of high necessity to assist with clinical decision-making and evaluate different treatment outcomes. Recently, the emergence of embedded intelligence has demonstrated great potential for continuous and real-time decision-making in various fields [16]. The integration of artificial intelligence and embedded hardware such as wearable devices enable increased operational efficiency, improved products and services, and enhanced performance. Indeed, driven by the advancement of sensing technology, several wearable sensors, e.g., inertial measurement units (IMUs), have been increasingly employed for convenient patient gait data collection due to the high wearability, reduced cost, and low power consumption [7]. The unprecedented data availability provides excellent opportunities for precise and efficient monitoring and diagnosis of ITW by using machine learning (ML) algorithms to mine the hidden gait pattern automatically. ML strategies for human movement recognition, such as neural networks, support vector machines (SVM), decision tree (DT), have been extensively investigated in a wide range of healthcare applications in the recent decade due to its ability to deal with high-dimensional and nonlinear data pattern, such as fall detection [17], Parkinson's disease stage prediction and severity assessment [18, 19], post-stroke patient gait pattern classification [20], human activity identification, e.g., walking versus running [21] and others [22-24]. Ilias et al. [25] proposed a combination of neural networks and SVM in classifying the gait patterns of autistic children from normal gait. They found that the SVM model with a polynomial kernel has the highest overall performance for classifying normal gait among children with autism [25]. Trentzsch et al. [26] tested six machine learning algorithms to differentiate gait between people with multiple sclerosis and healthy controls. They found the SVM model with Radial Basis Function (RBF) kernel could classify gait with the highest accuracy. Chakraborty et al. [27] investigated several regression modeling techniques to detect pathological gait. They found multiple adaptive regression splines (MARS) model (Accuracy=88.3%, Precision=0.89, Recall=0.87 and F1 Score=0.88) outperformed SVM (Accuracy=84.8%, Precision=0.85, Recall=0.83 and F1 Score=0.84) and logistic regression models (Accuracy=68.5%, Precision=0.78, Recall=0.51 and F1 Score=0.61) for gait classification. Pendharkar et al. [28] differentiated ITW gait patterns from normal ones, achieved accuracy of 87.5% using SVM based on heel accelerometry data collected from five children diagnosed with ITW and five normal healthy children. Pendharkar and coworkers [29] developed an automated way to assess the gait in children diagnosed with ITW through a threshold-value based statistical method using heel accelerometer data. The algorithm had an accuracy of 98.5%. However, the robustness of this algorithm is questionable due to changing frequency and speed during toe walking. Kim et al. [30] investigated the capabilities of ML algorithms to detect and differentiate heel-toe gait versus toe-toe gait using data from a single inertial sensor. They reported that k-means clustering was successful in differentiating toe-toe gait and heel-toe gait signals with 82% accuracy score.

This study's primary objective is to investigate the targeted ML approach to identify optimal sensor placements on the body for accurate identification of toe walking in children diagnosed with ITW. We hypothesize that sensors at the trunk level with statistical and frequency features would predict toe walking characteristics since trunk carries 2/3rd of body weight and represents movement of center of mass of body. Previous study has reported that foot posture and function affect trunk kinematics and lead to low back pain [31]. The secondary objective of this study is to investigate if data inputs as segmented gait cycles could improve the classification accuracy of these ML models. We plan to evaluate classifier performance using six conventional ML classifiers, including SVM, Random Forest (RF), DT, Multi-layer Perceptron (MLP) and Gaussian process (GP) to classify typical toe walking versus best heel strikes.

II. METHODS

A total of thirty-six children diagnosed with ITW (Age=9.4 ± 2.8 years, Height= 53.8 ± 6.6 cm, Weight=75.0 ± 27.2 lbs) participated in this study. All participants signed a written consent form before participation that Chapman Institutional Review Board (IRB) approved. Four wireless sensor modules composed of Xsens MTw sensors packaged in a 47mm×30mm×13mm plastic housing were used. The sensors contain 3D accelerometer and 3D rate gyroscopes to measure acceleration and angular velocities. The sensors weighed 16g, including the battery. The tri-axial accelerometer had a ±16g capacity in full range, and the gyroscope had ±2000°/s with a bandwidth of 3200Hz, where g represents acceleration due to gravity (1g =9.8 m/s^2). The accelerometer's sensitivity was 31.2 LSB/g and gyroscopes sensitivity were 14.375 LSB/s. The sensors were affixed at the posterior sacrum, posterior trunk at T4, left and right lateral shank just proximal to the lateral malleoli, Figure 1a shows the locations where the

sensors were affixed to the body and Figure 1b shows Xsens Awinda sensor unit.

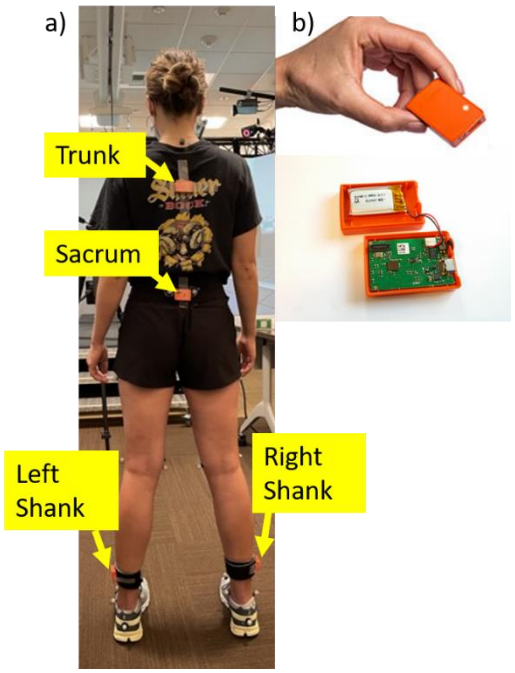
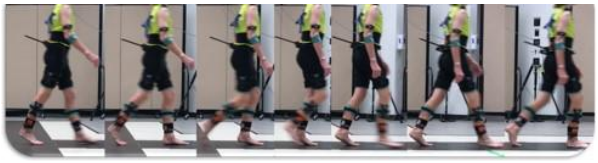


FIGURE 1. Representational picture a) participant with affixed sensors at Trunk, Sacrum, Right and Left Shanks, b) Xsens Awinda sensor unit.



a) Typical walking

b) i micking best heel strikes

FIGURE 2. a) Typical or toe walking , b) Best Heel Strikes (BHS) among children diagnosed with ITW

Data was collected using Xsens MT Manager Software suite. The sampling frequency was set to 75Hz. This is largely sufficient for human movement analysis in daily activities which occur in band-width [0.8-5 Hz][32].

The participants were instructed to i) walk with Best Heel Strikes (BHS), and ii) Toe Walk (TW) over a 15m long walkway as demonstrated in Figure 2. In a well-lit motion capture laboratory, all walking trials were conducted at the preferred walking speed. The walkway was embedded with two force plates (Bertec, Columbus, Ohio 43219). Typical gait in children with ITW may be affected due to white coat syndrome in the new testing environment, and children may consciously present BHS during walking in clinicians' presence. In an attempt to capture the child's typical toe walking, researchers asked parents if typical toe walking trials were similar to the gait observed at home. If parents reported it was not, researchers asked child to relax and verbally distracted the participant until the parent reported the gait was more similar to what they observed at home. We also instructed participants to mimic best heel strikes (BHS) with all participants trying their best to perform heel-to-toe gait. Each participant walked multiple trials of each walking type barefoot. Ten trials of 10 m walk were collected for each walk type (toe walking and Best Heel Strike). Each 10 m walk trial consisted of 3 to 4 gait cycles. Children with Idiopathic toe walking characteristics walk on their toes but have capability to make best heel strikes (BHS). Thus it is important to identify both toe walking and BHS among children. The gait remediation for these children would involve motor learning to produce BHS. In essence, this research is innovative in identifying toe walking in children using inertial sensors and could potentially be helpful in providing real-time corrective feedback to children.

If participants did not step at the center of the force plate or failed to perform the instructed kind of walk, the trial was repeated. Only complete walking trials of each category (typical and BHS walking) were used for the analysis. Figure 3 shows acceleration profiles during toe walking (TW) and best heel strike (BHS) gait for a) anterior posterior, b) medial lateral, c) vertical directions. The sensor was affixed at sacrum. The blue line represents BHS, the red line represents typical TW for children with ITW.

To investigate the efficacy of ML algorithms in differentiating TW versus BHS utilizing sensor signals, informative features were extracted from raw sensor data. Six ML algorithms, i.e., SVM, DT, RF, KNN, MLP and GP, were tested for gait classification. Sensor data input was fed to ML algorithms in two forms i) as a 10-second walking data and ii) truncated gait cycle signals (Figure 3). The effects of sensor node placement on gait classification accuracy was evaluated from i) one site and ii) all sites. Both classification scenarios used 10 s raw signals and truncated gait cycle signals, the best sensor placement analysis is also presented, providing a good reference for sensor placement prioritization.

A. FEATURE EXTRACTION

Feature extraction aims to extract informative features from raw signals, which can improve ML algorithms' accuracy. Commonly used methods for feature extraction from time-series signals are divided into two categories, statistical methods and transforming methods. The statistical methods, such as mean, standard deviation (SD), kurtosis, skewness etc., measure the fluctuation of signals and do not reflect the temporal characteristic of the signal, are the most common feature extraction approaches in signal processing-based applications.

Transforming methods aim at changing the signal into a different domain (frequency domain) and visualize the

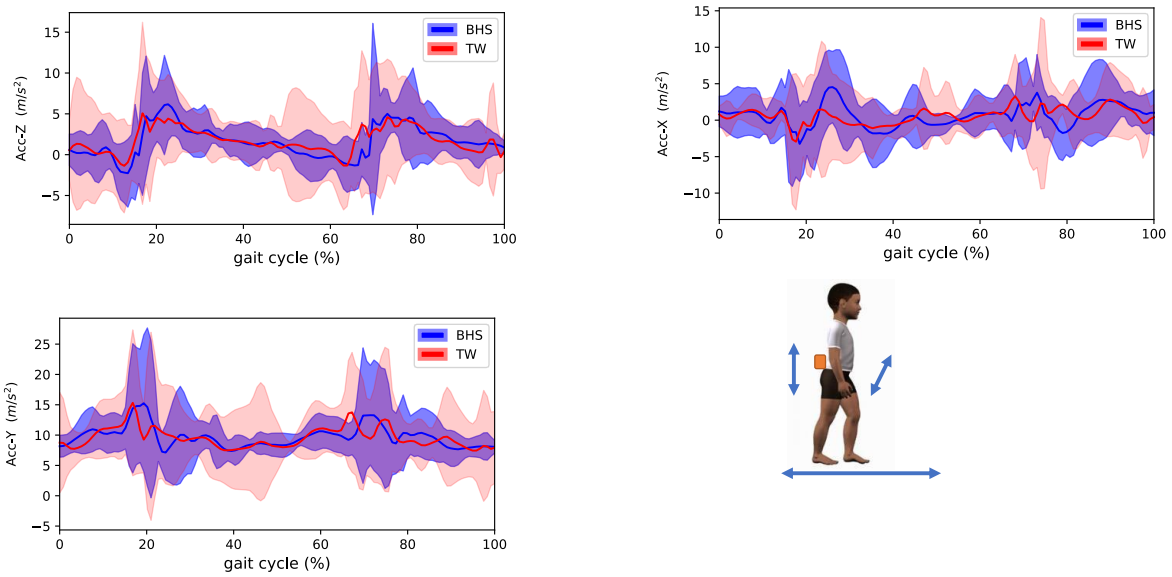


FIGURE 3. Acceleration profiles during best heel Strikes (BHS) and Toe walking (TW) presented by children with Idiopathic Toe Walking characteristics in a) anterior-posterior direction, b) medial-lateral direction, c) vertical direction, d) directions sensed by sensor. The shaded area represents standard deviation of acceleration curves for three trials of all subjects. The solid lines represent mean values of the acceleration during gait cycle. The blue solid line represents Best Heel Strike and red solid line represents toe walking trials.

TABLE 1 LIST OF TIME DOMAIN FEATURES EXTRACTED FROM THE ACCELERATION SIGNALS

Name of Feature	Brief Explanation	Formula
Mean	Average values of signal	$\frac{\sum_{t=1}^{T} x_t}{T}$
Max	The maximum value of the signal	$max(\boldsymbol{x})$
Min	The minimum value of the signal	$min(\boldsymbol{x})$
Standard deviation (SD)	Measure the dispersion of a signal	$\sqrt{\frac{\sum_{t=1}^{T}(x_t - \bar{x})^2}{T-1}}$
Skewness	The asymmetry of the probability density function of the signal	$\frac{\sum_{t=1}^{T}(x_t - \bar{x})^3}{(T-1) * SD^3}$
Kurtosis	The sharpness of the probability distribution of the signal	$\frac{\sum_{t=1}^{T}(x_t - \bar{x})^4}{(T-1) * SD^4}$
interquartile range (IQR)	The spread of the middle half of your distribution	$Q_3 - Q_1$
area under the curve of signal	the region bounded by the signal between the start and endpoints	-
area under the curve of the squared signal	the region bounded by the square of signal between the start and endpoints	-
median absolute deviation (MAD)	measure how spread out a set of data is	$median(\lvert x_t - \bar{x} \rvert)$

behaviors of data in that domain, i.e., discrete Fourier transform (DFT) to convert the signal to the frequency domain to characterize a signal with period/frequency, amplitude, phase. Assume the data collected at time t from each individual is denoted as $X_t = [x_{t,1}, x_{t,2}, \dots, x_{t,J}]$, where $j = 1,2, \dots, J$, J is the total number of signals (3D accelerometers and 3D gyroscopes). In this experiment, four sensors are affixed to each participant. Each 3D-accelerometer and 3D-gyroscope of each sensor measure along three orthogonal axes, X, Y, and Z. Thus $J = 24$ for four sensors and 6 channels of data. For each channel signal, both statistical and transformed features were extracted. The ten extracted statistical features in the time domain are listed in Table 1, where T is the length of channel data of each trial. Q_1 and Q_3 represent the first and third quantiles of the signal, respectively. The area under the curve was computed by integration using the trapezoidal rule [33].

Discrete Fourier transform (DFT) is applied to compute frequency domain features from the raw channel signals. DFT is a signal processing technique that transforms a signal into a vector of complex Fourier coefficients, which is defined by equation (1) below

$$X_k = \sum_{n=0}^{T-1} x_n e^{-\frac{2\pi i}{T} kn} \quad (1)$$

Where $0 \leq k \leq T-1$. The X_k represents the signal level at various frequencies. To facilitate the efficient computation of DFT, Fast Fourier Transform (FFT) is employed, which is an optimized algorithm for the implementation of DFT. The frequency components of a channel signal and its coefficients are determined using FFT. Figure 4 shows frequencies observed at the sacrum level during TW and BHS from accelerometers and gyroscope signals. It is clear that the TW

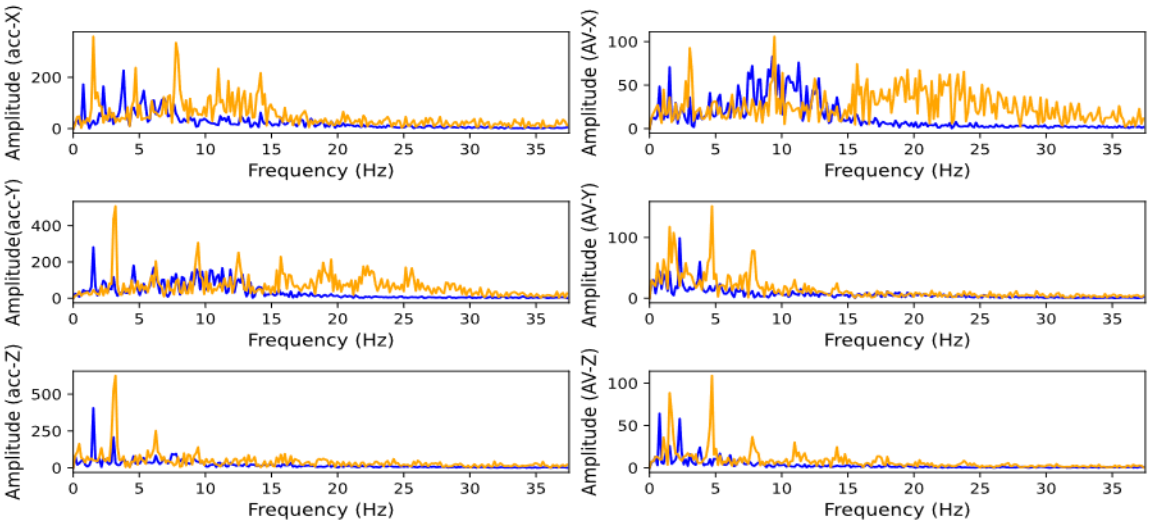


FIGURE 4. **Representations of the signals of sacrum in the frequency domain**

and BHS demonstrate obvious discriminative patterns. The frequency-domain features extracted from all accelerometers and gyroscope signals in this study include weighted mean frequency (1 feature), the first 5 DFT coefficients (5 features) and the first 5 maxima of DFT coefficients (5 features) and their corresponding frequencies (5 features). In addition, the statistical properties listed in Table 1 are also extracted from the frequency domain, which yields 26 features for each signal in total extracted from the frequency domain.

B. SEGMENTED GAIT CYCLES

Gait is a cyclic process and gait cycle is defined as the movement between two same gait events. For example, heel contact of one foot to next contact of the same foot. To generate a large number of data samples for better classification performance, the raw signals can be segmented into strides, that is, into gait cycles. The simplest way to identify the gait cycle from IMU signals is to detect the peak (a gait event) of a sensor signal, then the signal data between two consecutive similar peaks (similar gait events) can be considered as a gait cycle [34]. We have previously reported that gyroscope measurements in the sagittal plane are the best choice for gait segmentation because the measurements contain typical time-series patterns such as "valleys", "peaks"[35, 36]. Thus, we identified the local maximum as the segmentation points (Figure 5). Some local maxima may not correspond to gait cycles' actual starting or ending points. We enhanced the gait cycle detection method by constraining the minimum horizontal distance between neighboring peaks as 50 data points usual gait cycle times are around 1s). A '2 degrees/s' threshold could be set for gait cycle identification. In this way, some incorrect peaks can be disregarded. Figure 5 demonstrates the peak detection using the z-axis of the gyroscope from the shank sensor. Red stars indicate the detected peaks corresponding to a gait cycle's starting and ending points. The first and last peaks are the start and endpoints of a gait cycle. In other words, the signals are discarded before the first peak and after the last peak. Hence, gait initiation and gait termination data are carefully removed, leaving us with eight gait cycles (Figure 5). After the segmentation of the z-axis of gyroscope signals, the other signals (Acceleration-x, y, z, and gyroscope x, y) are segmented into gait cycles.

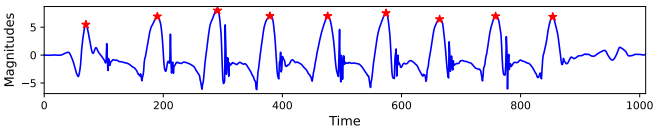


FIGURE 5. **Peak detection and truncation of raw sensor walking signals to segmented gait cycles (SGC)**

C. MACHINE LEARNING ALGORITHMS

Support Vector Machine (SVM): SVM transforms input data into a higher-dimensional space by a kernel function and then learns a boundary called hyperplane in that transformed space, which optimally separates data into two classes. SVM has gained wide popularity as a tool in pattern recognition and data classification due to its low computational cost, small memory occupation, and excellent performance in solving small samples and local extreme value problems [23, 37]. Given a set of N samples (x_i, y_i), i=1,2,…,N, where x_i is a vectors including multiple features, y_i is the true label for each sample i, the goal of SVM is to learn a classification function $f(x) = w^T x + b$, The solution of the problem is a vector of w and b that defines a separating hyperplane with the largest separation, or margin, between the two classes. The decision boundary can be found by minimizing the following constrained optimization problem [38].

$$loss = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$
$$\text{subject to } y_i(\boldsymbol{w}^T\boldsymbol{x} - b) \geq 1 - \xi_i, \text{ and } \xi_i \geq 0 \; for \; all \; i \quad (2)$$

where C is a tradeoff parameter between error and the margin. To generalize the linear decision classifier to nonlinear situations, the features need to map to a higher-dimensional space via some transformation $\emptyset: x \to \emptyset(x)$, then a kernel $K(x_i, x_j) = \emptyset^T(x)\emptyset(x)$ offers a more efficient and less expensive way to transform data into higher dimensions. Kernel functions commonly include linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.

Decision Tree (DT): DT is a non-parametric supervised technique that formulates the classification model in a tree structure for regression and classification [39]. The key idea of DT is to divide the dataset into smaller subsets into nodes and branches. A DT generally consists of one root, several branches, and many interval nodes. Every path is from the root node to a leaf node through the internal nodes. This path denotes a classification with the different conditions of the components. Every leaf node represents a response for regression or a class label for classification. A decision tree is constructed from the pre-classified data. The division of data into different classes is based on the values of the features of the given data. To determine which features to split, the commonly used method is to measure the impurity, which measures the homogeneity of the labels on a node. This process is applied to all subsets of data items recursively. The process terminates as all the data items in the current subgroup belong to the same class.

Random Forest (RF) [40]: RF is a variant of DT. Unlike DT, which builds a single tree on a whole dataset, RF creates a set of DTs using random resampling on the training set. For classification tasks, each DT then votes for a particular target class, and a class selected by most DTs is the output of the RF. RF benefits from two powerful techniques: bagging and random subspace selection. RF builds many DTs and allows each tree to randomly sample from the original dataset with replacement, resulting in different trees. Secondly, each tree picks a subset of features randomly. This forces more variation amongst the trees in the model, resulting in lower correlation and more diversification across trees. The hyperparameters, including the number of trees (n_estimators) and the number of variables (max_features), must be optimized to improve the classification accuracy.

K-Nearest Neighbors (KNN) [41]: KNN tries to predict outputs by calculating the distance between the test data and training points, then selecting the K number of points closest to the test data. A class label is assigned based on a majority vote for classification problems. A reasonable distance function will identify essential features and discriminate between relevant and irrelevant ones. Commonly used distance functions include Euclidean, Manhattan, and Hamming.

Euclidean distance is calculated as the square root of the sum of the squared differences between a new point and an existing point. It is the minimum distance between two points. If points (x_1, y_1) and (x_2, y_2) are in 2-dimensional space, then the Euclidean distance 'd' is represented as

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

Manhattan distance between two points a and b with k dimensions is defined as 'D' below

$$D = \sum_{j=1}^{k} |a_j - b_j| \quad (4)$$

Hamming distance is mainly used for categorical variables, also referred as binary strings. For example, the two points, (0,1,1) and (0,1,0), the hamming distance is 1, since only one value (last value) is different between two variables.

Multi-layer Perceptron (MLP): Multilayer perceptron (MLP) is one of the most commonly used types of artificial neural networks. The standard architecture of an MLP artificial neural network consists of an input layer, multiple hidden layers, and an output layer. Hidden layers learn representations of input data by using non-linearity functions. The number of neurons in the input layer equals the number of features, and the number of neurons in the output layer equals the number of classes. In contrast, the number of neurons in each hidden layer needs to be tuned to find a suitable network with sufficient parameters and good generalization for classification or regression tasks.

Gaussian process (GP): A GP is a stochastic process whose kernel is a Gaussian distribution [42]. The GP assumes that the mapping from inputs and outputs via a latent function f, which can be defined mathematically as

$$f \sim GP(m(x), k(x, x')) \quad (5)$$

Where $m(x)$ and $k(x, x')$ are the mean and covariance functions respectively, denoted by

$$m(x) = E(f(x))$$
$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))^T] \quad (6)$$

The output values are assumed to be independent when conditioned on the latent function, i.e., $p(y|\boldsymbol{x}, f) = \prod_{i=1}^{N} p(y_i|f(\boldsymbol{x}_i))$. Kernel function $k(x, x')$ is the critical ingredient in using Gaussian processes, which determine the shape of prior and posterior of the GP. The main advantage of GP is probabilistic so that one can compute empirical confidence intervals to quantify the uncertainty of the prediction. When applying it to classification tasks, the posterior of the latent function f no longer has a closed-form solution since a Gaussian likelihood is inappropriate for discrete class labels. Several approximation schemes have been suggested, including Laplace's method, variational approximations, mean-field methods, Markov chain Monte Carlo and Expectation Propagation.

Deployment of ML algorithms: All machine learning codes were deployed using Python, and computations were performed on a 2.3 GHz Quad-Core Intel Core i7 processor. The classification accuracy was determined for i) unsegmented 10 m walk dataset and ii) segmented gait cycles (SGC). Four hundred ninety-two walking samples are obtained from the experiment. The whole data is randomly shuffled and split into training and test with a ratio of 80%: 20%, which corresponds to 393 and 99 samples, respectively. For each sample, 36 features are extracted for each signal, i.e.,

corresponding to 36*24=864 features per sample. Once we have extracted features, min-max scaling is used to normalize the data to range between 0 and 1 to reduce variation. Five-fold cross-validation is used on the training data set for each classifier to select the best model for the test. We used a grid search algorithm to optimize hyperparameters when performing five-fold cross-validation for each classifier; this tuning technique exhaustively generates candidates from a grid of parameter values, then builds a model for every combination of hyperparameters specified and evaluates the accuracy of each model.

Hyperparameter Tuning: We conducted a grid search for each ML classifier, four kernels of SVM, including linear, polynomial, radial basis function (RBF), and sigmoid with a set of regularization parameter $C \in [0.01, 0.05, 0.1, 1, 10]$ are tuned. For the DT, two attribute selection methods, entropy to calculate information gain and 'Gini index' for the Gini impurity, are tuned. For the RF, we tune a wide range of values from 10 to 700 for the number of trees (n_estimators) in the forest. For the KNN, grid search is performed over various values of K. For the MLP, different numbers and sizes of hidden layers, activation functions including Rectified Linear unit (ReLU), tanh, logistic, and diverse learning rates are employed. For the GP the best kernel is chosen from RBF, DotProduct, Matern, RationalQuadratic, WhiteKernal. The model with the highest validation accuracy is eventually selected for the test. The best hyperparameters configuration for each classifier is listed in

Table 2. Classification performance was evaluated using several performance metrics such as accuracy, precision, sensitivity, specificity, and the Matthews correlation coefficient (MCC). Accuracy is the ratio of correctly identified samples out of all predictions. Sensitivity, or recall, also known as true positive rate (TPR), refers to the proportion of true positives to actual total positive predictions. It is a measure of how well a model can identify true positives. Specificity, or true negative rate (TNR), on the other hand, refers to the ratio of true negatives to total negatives in the data. Precision is the ratio of true positives to the combined number of true positives and false positives, which measures the model's accuracy in classifying a sample as positive. F1-score is the harmonic mean of the precision and recall. The Matthews correlation coefficient (MCC) measures the overall association between actual classes and predicted classes by calculating the correlation coefficient. It is, in essence, a correlation coefficient value between -1 and +1. A coefficient of +1 represents a perfect prediction. Conversely, -1 represents the worst prediction where a classifier labels all the positives as negatives and all the negatives as positives. 0 indicates an average random prediction. More concretely, the metrics are calculated using the following formulas:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (7)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (8)$$

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$F1 - score = \frac{2TP}{2TP + FN + FP} \quad (11)$$

TP, FN, FP, and *TN* represent true positive, false negative, false positive, and true negative. *TP* and *TN* imply the number of the positive/negative classes that have been correctly categorized; *FN* (false negative) are the reverses, indicating the number of positive/negative samples categorized wrong.

TABLE 2 BEST HYPERPARAMETERS CONFIGURATIONS FOR DIFFERENT ML ALGORITHMS USED FOR CLASSIFICATION OF TOE WALKING (TW) VERSUS BEST HEEL STRIKE (BHS)

Model	Best Parameter Configuration
SVM	C: 0.05, kernel:poly, degree:3
DT	criterion: gini, max_depth: 18, min_samples_leaf: 10
RF	bootstrap: True, max_depth: 20, max_features: auto, n_estimators: 600
KNN	leaf_size: 20, metric: minkowski, n_neighbors: 2, p: 1, weights: distance
MLP	activation: relu, alpha: 0.0001, hidden_layer_sizes: (50, 10), learning_rate: invscaling, solver: adam
GP	kernel: 1**2 * Matern (length_scale=1, υ=1.5)

III. RESULTS

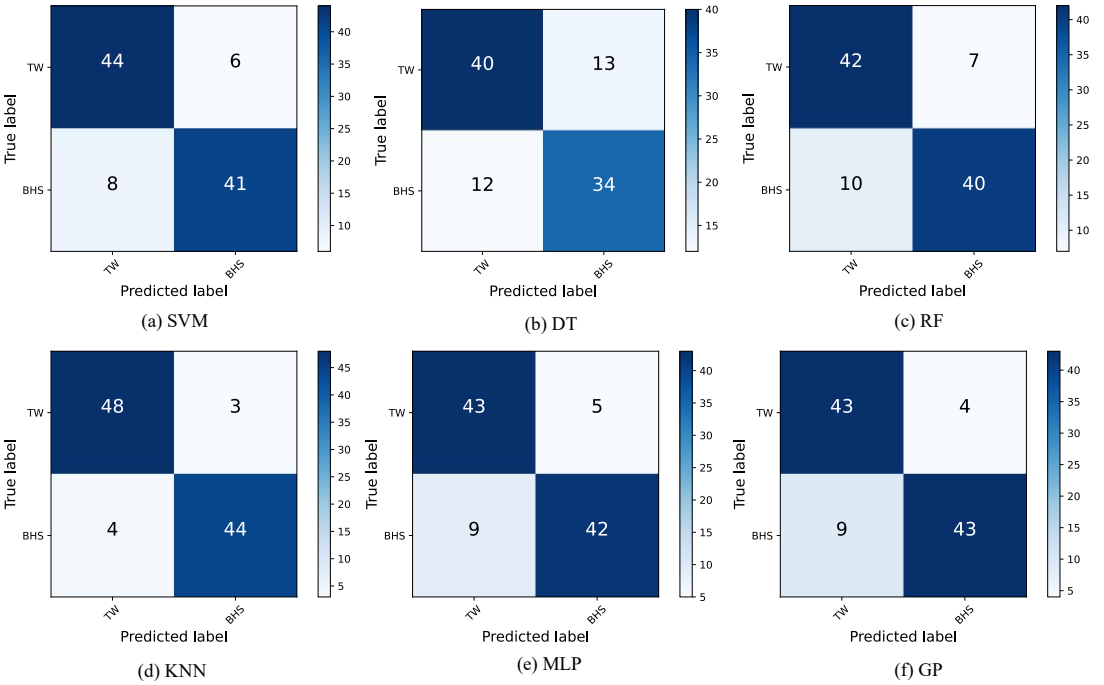
TABLE 3 PERFORMANCE METRICS OF THE CLASSIFICATION MODELS WITH ALL FOUR BODY ATTACHED SENSORS. THE HIGHEST PERFORMING CLASSIFIER IS HIGHLIGHTED IN BOLD.

Model	Performance metrics					
	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
SVM	0.8585	0.8800	0.8461	0.8723	0.8627	0.7176
DT	0.7474	0.7547	0.7692	0.7234	0.7619	0.4932
RF	0.8282	0.8571	0.8076	0.8510	0.8316	0.6579
KNN	**0.9292**	**0.9411**	**0.9230**	**0.9361**	**0.9320**	**0.8585**
MLP	0.8585	0.8958	0.8269	0.8936	0.8600	0.7199
GP	0.8686	0.9148	0.8269	0.9148	0.8686	0.7418

Classification capabilities of six different classifiers to distinguish the toe walking (TW) and best heel strike (BHS) was evaluated. We found all of these six algorithms can classify BHS and TW gait patterns from the data correctly with an average prediction accuracy of 84.84%. We found KNN showed the best classification accuracy of 92.92%, specificity of 93.61% and sensitivity of 92.30%, F1-score of 93.20%, MCC of 0.8585 to distinguish between TW and BHS, which is significantly higher (6.9%) than GP (Table 3 and Figure 6). Among 99 test samples, only 3 BHS and 4 TW samples are misclassified. We found GP achieves satisfiable performance with an accuracy of 86.86% when using Matern

kernel with smooth parameter ν=1.5. The atern kernel is a generalization of the RBF. Compared to RBF, Matern has an

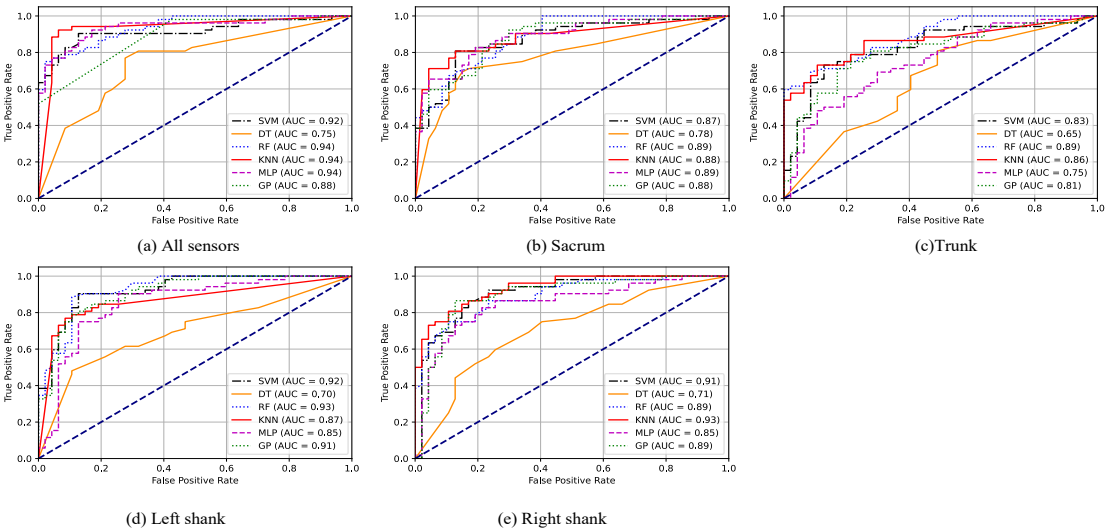
This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3192136



(a) SVM (b) DT (c) RF

(d) KNN (e) MLP (f) GP

FIGURE 6. Confusion Matrix using all four body affixed sensors



(a) All sensors (b) Sacrum (c)Trunk

(d) Left shank (e) Right shank

FIGURE 7. ROC curves and AUC values are presented for a) all four sensors, b) sacrum sensor, c) Trunk sensor, d) left shank sensor, and e) right shank sensor for different ML algorithms.

additional parameter ν to control the smoothness of the estimated function. The smaller ν, the less smooth the function gets. When ν→∞, the kernel gets equivalent to RBF. LP with 2-hidden-layers and SVM with a polynomial kernel achieve the same accuracy of 85.85% with 14 samples misclassified.

Nevertheless, the specificity of MLP is slightly higher than that of SVM. MLP and SVM perform better than RF and far better than DT in toe walking classification accuracy. We found that for all the classifiers except DT, the specificity values are higher than those of sensitivity, indicating that the BHS samples are more rarely misclassified than the Toe walking samples.

Table 4 shows the performance results with four different sensor locations, and the best performing metrics are highlighted in bold. The comparison of receiver operating characteristic (ROC) curves with varying sensor locations using six classifiers in Figure 7. The ROC curve is created by plotting the TPR against the FPR at various threshold settings, which is a typical plot to characterize the diagnostic ability [43]. The area under the ROC curve (AUC) is also demonstrated in Figure 7.

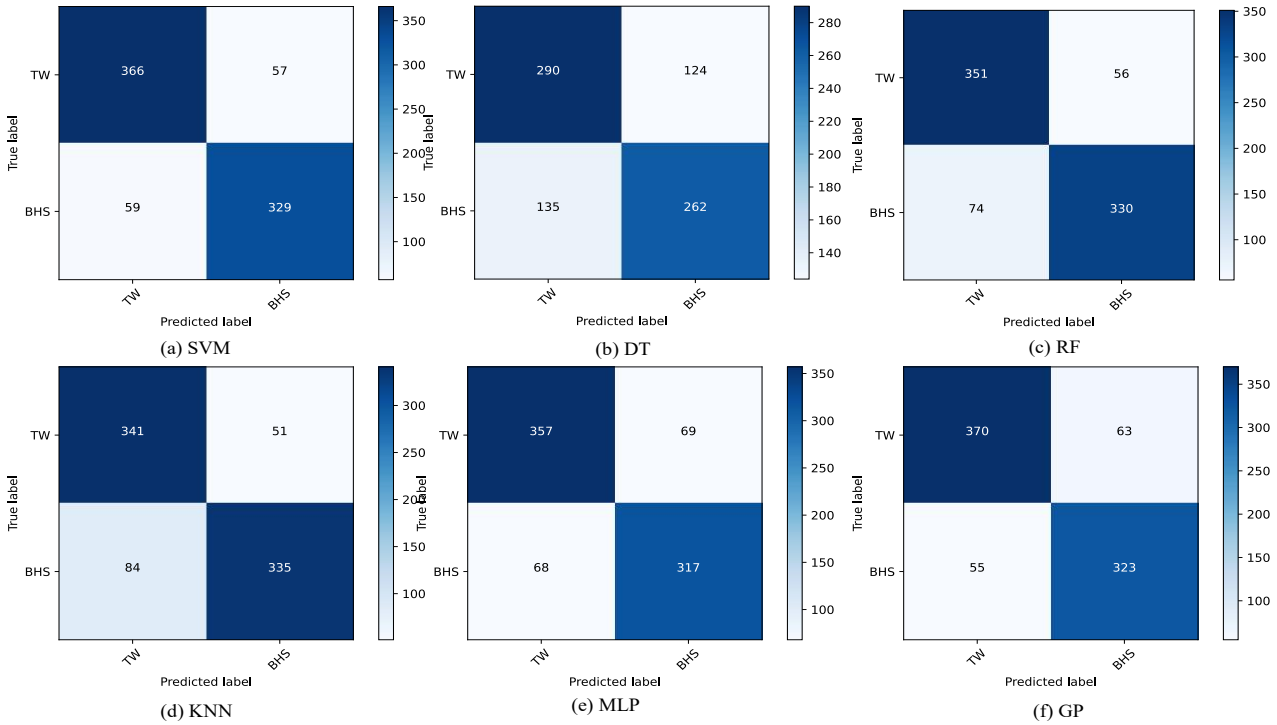


FIGURE 8. Confusion matrix for the six ML classifiers using the gait cycle segmented data of all sensors.

TABLE 4 PERFORMANCE METRICS OF THE CLASSIFICATION MODELS WITH DIFFERENT SENSOR LOCATIONS

Sensor Location	Model	Performance					
		Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
Sacrum	SVM	0.7878	0.7818	**0.8269**	0.7446	0.8037	0.5744
	DT	0.7692	**0.8222**	0.7115	**0.8297**	0.7628	0.5428
	RF	0.7575	0.7692	0.7692	0.7446	0.7692	0.5139
	KNN	**0.7979**	0.8076	0.8076	0.7872	**0.8076**	0.5949
	MLP	**0.7979**	0.8200	0.7884	0.8085	0.8039	**0.5962**
	GP	0.7676	0.7959	0.7500	0.7872	0.7722	0.5365
Trunk	SVM	**0.7878**	0.8444	0.7307	0.8510	**0.7835**	**0.5835**
	DT	0.6363	0.6481	0.6730	0.5957	0.6603	0.2695
	RF	0.7575	0.8043	0.7115	0.8085	0.7551	0.5206
	KNN	0.7777	**0.8750**	0.6730	**0.8936**	0.7608	0.5766
	MLP	0.6666	0.6727	0.7115	0.6170	0.6915	0.3301
	GP	0.7474	0.7647	**0.7500**	0.7446	0.7572	0.4942
Left shank	SVM	**0.8585**	**0.8958**	0.8269	**0.8936**	0.8600	**0.7199**
	DT	0.6464	0.6808	0.6153	0.6464	0.6464	0.2962
	RF	**0.8585**	**0.8958**	0.8269	**0.8936**	**0.8686**	**0.7199**
	KNN	0.8181	0.8541	0.7884	0.8510	0.8200	0.6390
	MLP	0.7777	0.8000	0.7692	0.7872	0.7843	0.5557
	GP	0.8282	0.8301	**0.8461**	0.8085	0.8380	0.6554
Right shank	SVM	0.8282	0.8070	**0.8846**	0.7659	0.8440	0.6573
	DT	0.6666	0.7209	0.5961	0.7446	0.6526	0.3433
	RF	0.7777	0.8125	0.7500	0.8085	0.7800	0.5580
	KNN	0.8383	0.8214	**0.8846**	0.7872	0.8518	0.6768
	MLP	0.7979	0.7758	0.8653	0.7234	0.8181	0.5969
	GP	**0.8686**	**0.8823**	0.8653	**0.8723**	**0.8737**	**0.7371**

The AUC summarizes the classification quality and is a measure of accuracy, where an AUC of 0.5 indicates a random classifier with no value. The best accuracy for all six classifiers is obtained using all four sensors for each algorithm. The maximum AUC value (0.94) is achieved from KNN, MLP, and RF when applied the algorithms to the data that combines all sensors. Another phenomenon is that none of the algorithms consistently outperformed the others in terms of accuracy. For example, SVM performs best at trunk location but does not perform well than GP when using the suitable shank sensor. We found performance at left and right shank locations is better than at sacrum and trunk. Sensor signal segmentation into gait cycles resulted in a large number of samples.

After the segmentation into gait cycles, the data size increases to 4044 (3233 for training and 811 for the test), close to 10 times the original size. The classification results using the same feature extraction strategy were evaluated. Figure 8 shows the confusion matrix for the six classifiers using the gait cycle segmented data. The quantitative results are provided in Table 5. We found SVM had an accuracy of 85.69%. However, KNN yielded the best precision and specificity but relatively low sensitivity. GP demonstrated the best sensitivity, indicating the best diagnostic ability to identify the TW patterns. Similarly, the DT showed poor ability to distinguish the TW and BHS and had the lowest accuracy of 68.06% and smallest MCC value of 0.3607. The performance is still promising despite the lower average accuracy of 81.60% and F1-score of 82.23% than that of using raw data. The effects of the number of sensors and sensor placement are explored, and the ROC curves and classification performance are provided in Figure 9 and Table 5. The results show that using all sensors gains the best classification results, implying that richer information leads to better discriminative power.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3192136

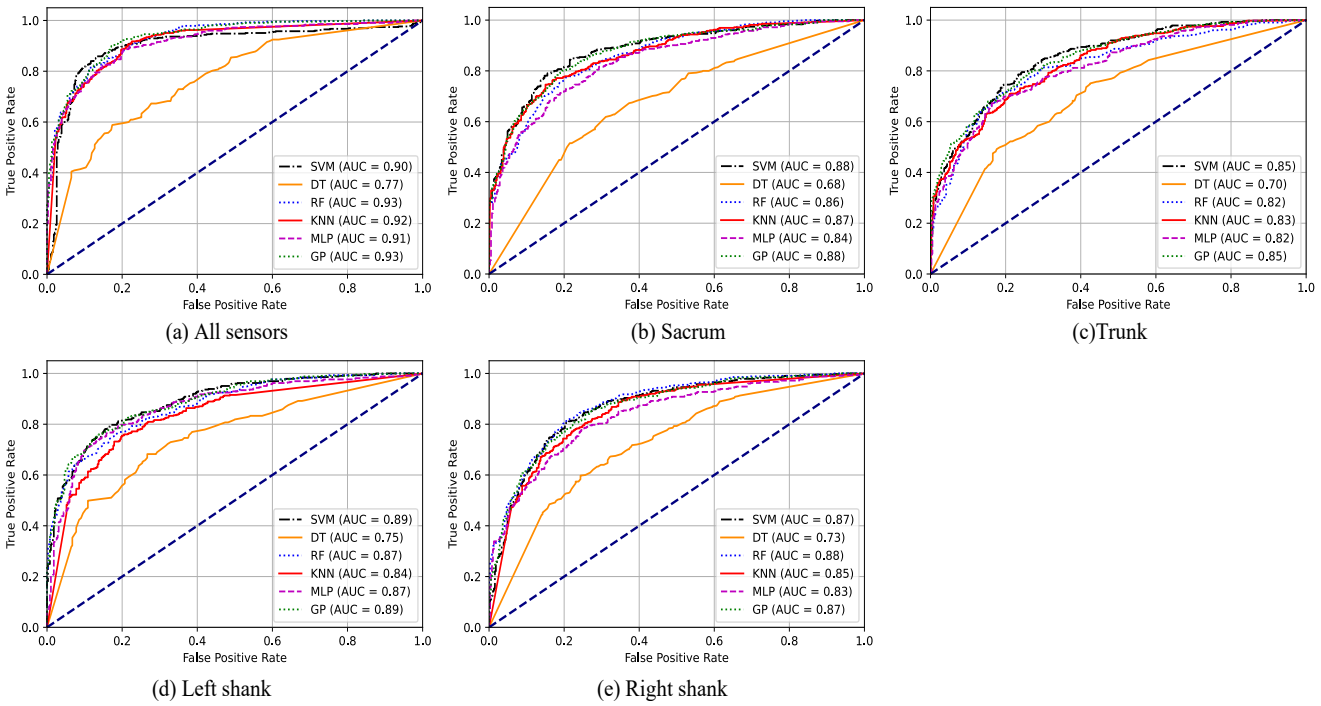


FIGURE 9. ROC curves and AUC values are presented for a) all four sensors, b) sacrum sensor, c) Trunk sensor, d) left shank sensor, and e) right shank sensor for different ML algorithms.

TABLE 5 PERFORMANCE METRICS OF THE CLASSIFICATION MODELS USING SEGMENTED DATA

Sensor Location	Model	Performance					
		Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
All Sensors	SVM	**0.8569**	0.8652	0.8611	0.8523	**0.8632**	**0.7133**
	DT	0.6806	0.7004	0.6823	0.6787	0.6912	0.3607
	RF	0.8397	0.8624	0.8258	0.8549	0.8437	0.6800
	KNN	0.8335	**0.8698**	0.8023	**0.8678**	0.8347	0.6698
	MLP	0.8310	0.8380	0.8400	0.8212	0.8390	0.6613
	GP	0.8545	0.8545	**0.8705**	0.8367	0.8624	0.7081
Sacrum	SVM	**0.8076**	**0.8225**	**0.8070**	**0.8082**	**0.8147**	**0.6148**
	DT	0.6461	0.6546	0.6823	0.6062	0.6689	0.2894
	RF	0.7718	0.7803	0.7858	0.7564	0.7831	0.5425
	KNN	0.7903	0.8179	0.7717	0.8108	0.7941	0.5820
	MLP	0.7558	0.7609	0.7788	0.7305	0.7697	0.5101
	GP	0.8027	0.8192	0.8000	0.8056	0.8095	0.6051
Trunk	SVM	**0.7644**	**0.7896**	**0.7505**	0.7797	**0.7696**	**0.5297**
	DT	0.6461	0.6691	0.6423	0.6502	0.6554	0.2922
	RF	0.7533	0.7892	0.7223	**0.7875**	0.7542	0.5097
	KNN	0.7459	0.7717	0.7317	0.7616	0.7512	0.4928
	MLP	0.7287	0.7506	0.7223	0.7357	0.7362	0.4575
	GP	0.7570	0.7864	0.7364	0.7797	0.7606	0.5157
Left shank	SVM	**0.8014**	0.8055	**0.8188**	0.7823	**0.8121**	0.6017
	DT	0.7003	0.7136	0.7152	0.6839	0.7144	0.3992
	RF	0.7817	0.7966	0.7835	0.7797	0.7900	0.5629
	KNN	0.7755	0.8060	0.7529	0.8005	0.7785	0.5529
	MLP	0.7965	**0.8201**	0.7835	**0.8108**	0.8014	0.5937
	GP	**0.8014**	0.8142	0.8047	0.7979	0.8094	0.6023
Right shank	SVM	0.7940	0.7905	0.8258	07590	0.8078	0.5869
	DT	0.6744	0.6968	0.6705	0.6787	0.6834	0.3489
	RF	**0.8027**	**0.7977**	**0.8352**	0.7668	**0.8160**	**0.6043**
	KNN	0.7731	0.7822	0.7858	0.7590	0.7840	0.5450
	MLP	0.7657	0.7701	0.7882	0.7409	0.7790	0.5299
	GP	0.7903	0.7972	0.8047	**0.7746**	0.8009	0.5796

This is consistent with classification results obtained using raw signals. Figure 9 shows that except DT all other 5 classifiers are comparable and could be good candidates for the ITW monitoring system. In terms of the location of sensors, configuring all four sensors together always performed the best. Power spectral density evaluations were conducted for accelerations in anterior posterior, medial-lateral and vertical directions. Power in four frequency bands was computed i) 0 Hz < f ≤ 2 Hz, ii) i) 2 Hz < f ≤ 4 Hz, iii) 4 Hz < f ≤ 6 Hz, and iv) 6 Hz < f ≤ 8Hz Figure 10). The dominant frequencies for BHS and TW are presented in Figure 11.

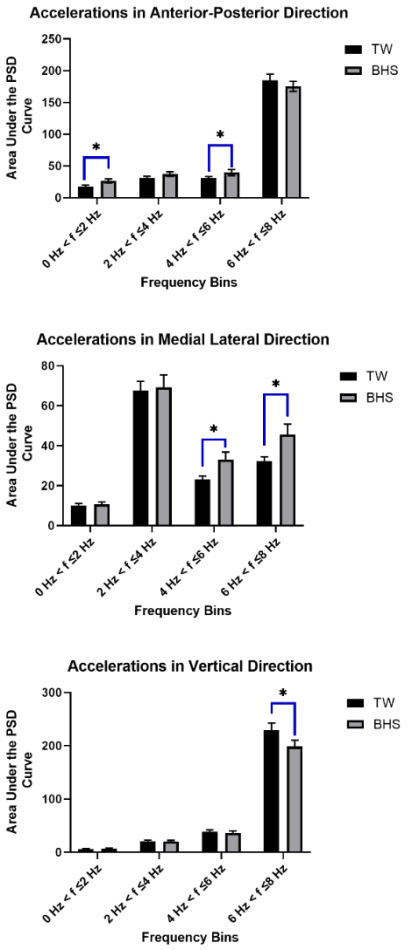


FIGURE 10. Area under the power spectral density (PSD) curves were computed for 4 frequency bands i) 0Hz < f ≤ 2Hz, ii) 2Hz < f ≤ 4 Hz, iii) 4Hz < f ≤ 6 Hz, iv) 6 Hz < f ≤ 8Hz for all three directions of accelerations.

IV. DISCUSSION

This study aimed to determine optimal sensor location and ML algorithms that can classify toe walking among children diagnosed with ITW with high accuracy. Previous studies have used the accelerometer to differentiate the toe-walking stance from the normal stance [14, 44-46] among children with ITW. In this study, we demonstrated the influence of i) 15 m walk continuous timeseries versus SGC, ii) body sensor locations (right and left shank, trunk, and sacrum), and iii) various ML algorithms (SVM, DT, RF, KNN, MLP and GP) in accurately classifying toe walking gait. This research provides a platform for ML-based automated classification of toe walking needed for intervention among ITW children. The highest classification of 92.9% was observed using all four sensors using KNN. This was followed by GP (86.8%), SVM and MLP (85.8%), RF (82.8%) and DT (74.7%).

ITW participants demonstrated reduced variability when mimicking BHS, as shown in Fig. 3(a-c). Gait variability is associated with the energy cost of walking in people with multiple sclerosis [47]. Thus, more variability found during toe walking could increase walking energetics [48-51]. Lower limb muscle fatigue is more commonly found in children who walk than in typically developing peers [52, 53].

We evaluated the effects of input timeseries lengths for ML classification. We utilized a simple yet effective peak detection method similar to our previous study to truncate raw signals to gait cycles [36]. We tested the influence of input timeseries by employing i) 15 m long walking timeseries signals versus SGC. Although the number of samples increased, classification accuracy was not affected due to the truncation of time series into gait cycles. We found increased number of samples reduced overfitting of the classifier and boosted classification performance.

We used both temporal and frequency domain features (Table 1). The frequency components of accelerations at the sacrum were analyzed at different frequency bins (0-2 Hz, 2-4Hz, 4-6Hz, and 6-8Hz). Paired t-test was carried out to compare the median frequencies between BHS and TW. We found BHS resulted in significantly higher frequencies in medial-lateral (ML) and anterior-posterior (AP) directions, whereas TW resulted in significantly higher accelerations in the vertical direction (Figure 10). Specifically, vertical accelerations produced during TW in the sacrum consist mainly of high frequencies. TW could potentially lead to low back pain among children diagnosed with ITW[54]. The dominant frequency during TW was higher than BHS (28Hz versus 26.6) (Figure 11).

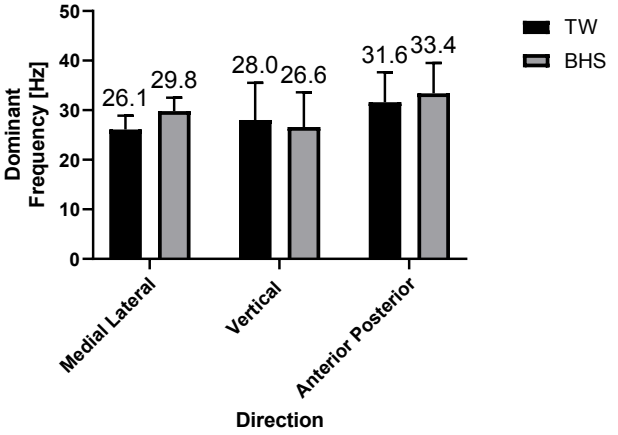


FIGURE 11. Mean dominant acceleration frequencies during TW and BHS.

Six ML algorithms are tested, and hyperparameter configurations are tabulated in Table 2. To avoid the overfitting of ML models, we used five-fold cross-validation techniques. The confusion matrix using all four sensors among six ML algorithms is provided in Figure 6. The accuracy of all six classifiers was compared using Receiver Operating Characteristics (ROC) curves (Figure 7), confusion matrices, and other performance metrics like Mathews Correlation Coefficient (MCC) (Table 5). We found KNN classified the TW patterns with highest accuracy (92.9%), highest precision (94.1%), sensitivity (92.3%), specificity (93.6%) with F1 score (93.2%) and MCC (85.8%).

We evaluated the effects of body sensor locations on the accuracy of toe gait classification. Four sensors affixed at both shanks, trunk, and sacrum. To perform comparisons, same hyperparameter tuning methods were conducted. We found a single sensor located at the shank could classify TW versus BHS with an accuracy of 86.8%, followed by sacrum (79.7%) and trunk (78.7%). Our results demonstrate that sensors located at the shank have added advantages for accurate gait classification. It may be attributed to knee and hip joints that may absorb or produce counteractive movements peculiar to gait type, thus reducing classification accuracy. We thus observed lower accuracies at sacrum and trunk positions. The ROC curves and AUC values are shown in Figure 9.

There is insufficient research on whether existing clinical interventions are adequate ITW treatment options. We propose that if TW is accurately identified using sensors affixed at the shank level, appropriate real-time feedback interventions can be implemented. Gait classification accuracy may be affected by intra-subject variability and severity of toe walking. Thus, the heterogeneity in toe walking severity among children diagnosed with ITW may challenge ML algorithms to classify gait accurately. In addition, the effects of an environment may influence walking behavior in children. For example, in the presence of an experimenter/clinician, the children may present the best performance to maintain good foot contact. Participants were asked to look forward at the target (20 m far) while walking in laboratory settings. The consistency was held in every trial during data collection.

Since the classification accuracy is also affected by the sample size of the data. If the data size is small, even a few misclassified samples will reduce the performance obviously from a statistical perspective. However, a higher classification performance can be achieved if more datasets are available. Nevertheless, the ML algorithms, especially SVM, KNN, and GP, have good discriminative power for gait classification of ITW children and can potentially be integrated into an expert gait system for monitoring and diagnosis. We found that all ML algorithms showed good specificity, which indicates that the ML algorithms can accurately identify BHS patterns, so the number of false positives is low, and this is helpful from the treatment perspective. The children will be less interrupted by the false alarm. Finally, this study provides evidence that an ML-enabled low-cost gait monitoring device can give a good capability for monitoring without hindrance.

V. CONCLUSION

There is limited research on using automated algorithms to identify TW in children diagnosed with ITW using wearable sensors and providing real-time feedback for correcting gait. In this study, we evaluated the performance of wearable sensors located at four different body locations using six ML algorithms to classify TW. We found ML algorithms successfully classified the TW strides from BHS in the children with ITW. This study demonstrates the significant potential of using low-cost wearable devices and ML algorithms to diagnose and monitor the gait of ITW children and further intervene using feedback. Accurate quantification of toe walking steps is critical for designing new real-time interventions for children diagnosed with ITW. Using wearable sensors and ML, real-time TW stride detection can be integrated with closed-loop control in assistive devices for intervention and motor rehabilitation. This will reduce costs and the burden on both clinicians and parents of the children. Future work will include additional participants and model personalization to improve performance.

ACKNOWLEDGMENT

We are thankful to Richard Beuttler, Michael Shiraishi, and DPT students during data collection.

REFERENCES

[1] A. J. Caserta, V. Pacey, M. C. Fahey, K. Gray, R. H. Engelbert, and C. M. Williams, "Interventions for idiopathic toe walking," *Cochrane Database of Systematic Reviews,* no. 10, 2019.

[2] R. Soangra *et al.*, "Foot Contact Dynamics and Fall Risk among Children Diagnosed with Idiopathic Toe Walking," *Applied Sciences,* vol. 11, no. 6, 2021, doi: 10.3390/app11062862.

[3] J. J. Ruzbarsky, D. Scher, and E. Dodwell, "Toe walking: causes, epidemiology, assessment, and treatment," *Curr Opin Pediatr,* vol. 28, no. 1, pp. 40-6, Feb 2016, doi: 10.1097/MOP.0000000000000302.

[4] V. De Oliveira, L. Arrebola, P. De Oliveira, and L. Yi, "Investigation of Muscle Strength, Motor Coordination and Balance in Children with Idiopathic Toe Walking: A Case-control Study," *Developmental Neurorehabilitation,* vol. 24, no. 8, pp. 540-546, 2021, doi: 10.1080/17518423.2021.1899326.

[5] C. M. Williams, P. Tinley, and M. Curtin, "The Toe Walking Tool: a novel method for assessing idiopathic toe walking children," *Gait & posture,* vol. 32, no. 4, pp. 508-511, 2010.

[6] A. A. van Kuijk, R. Kosters, M. Vugts, and A. C. Geurts, "Treatment for idiopathic toe walking: a systematic review of the literature," *Journal of rehabilitation medicine,* vol. 46, no. 10, pp. 945-957, 2014.

[7] R. O'Sullivan, K. unir, and L. Keating, "Idiopathic toe walking—A follow-up survey of gait analysis assessment," *Gait & posture,* vol. 68, pp. 300-304, 2019.

[8] R. Engelbert, J. W. Gorter, C. Uiterwaal, E. van de Putte, and P. Helders, "Idiopathic toe-walking in children, adolescents and young adults: a matter of local or generalised stiffness?," *BMC Musculoskelet Disord,* vol. 12, p. 61, Mar 21 2011, doi: 10.1186/1471-2474-12-61.

[9] R. O'Sullivan, K. Munir, and L. Keating, "Idiopathic toe walking-A follow-up survey of gait analysis assessment," *Gait Posture,* vol. 68, pp. 300-304, Feb 2019, doi: 10.1016/j.gaitpost.2018.12.011.

[10] P. Engstrom and K. Tedroff, "Idiopathic Toe-Walking: Prevalence and Natural History from Birth to Ten Years of Age," *J Bone Joint Surg Am,* vol. 100, no. 8, pp. 640-647, Apr 18 2018, doi: 10.2106/JBJS.17.00851.

[11] S. Armand, E. Watelain, M. Mercier, G. Lensel, and F. X. Lepoutre, "Identification and classification of toe-walkers based on ankle kinematics, using a data-mining method," *Gait Posture,* vol. 23, no. 2, pp. 240-8, Feb 2006, doi: 10.1016/j.gaitpost.2005.02.007.

[12] S. S. Bidabadi, I. Murray, G. Y. F. Lee, S. Morris, and T. Tan, "Classification of foot drop gait characteristic due to lumbar radiculopathy using machine learning algorithms," *Gait & posture,* vol. 71, pp. 234-240, 2019.

[13] M. Pollind, R. Soangra, M. Grant-Beuttler, and A. Aminian, "Customized Wearable Sensor-Based Insoles for Gait Re-Training in Idiopathic Toe Walkers," *Biomed Sci Instrum,* vol. 55, no. 2, pp. 192-198, Apr 2019.

[14] S. Kim, R. Soangra, M. Grant-Beuttler, and A. Aminian, "Wearable Sensor-Based Gait Classification in Idiopathic Toe Walking Adolescents," *Biomed Sci Instrum,* vol. 55, no. 2, pp. 178-185, Apr 2019.

[15] G. Ershadi, M. Gwak, A. Aminian, R. Soangra, M. Grant-Beuttler, and M. Sarrafzadeh, "Smart Insole: Remote Gait Detection Algorithm Using Pressure Sensors For Toe Walking Rehabilitation," presented at the 2021 IEEE 7th World Forum on Internet of Things (WF-IoT), 2021.

[16] K. P. Seng and L.-m. Ang, "Embedded Intelligence: State-of-the-Art and Research Challenges," *IEEE Access,* 2022.

[17] H. Sadreazami, M. Bolic, and S. Rajan, "Contactless fall detection using time-frequency analysis and convolutional neural networks," *IEEE Transactions on Industrial Informatics,* 2021.

[18] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease," *Applied Soft Computing,* vol. 94, p. 106494, 2020.

[19] F. a. B. R. K. a. H. C. J. a. H. S. a. A. D. C. Wahid, "Classification of Parkinson's Disease Gait Using Spatial-Temporal Gait Features," *IEEE Journal of Biomedical and Health Informatics,* vol. 19, no. 6, pp. 1794-1802, 2015, doi: 10.1109/JBHI.2015.2450232.

[20] K. Kaczmarczyk, A. Wit, M. Krawczyk, and J. Zaborski, "Gait classification in post-stroke patients using artificial neural networks," *Gait & posture,* vol. 30, no. 2, pp. 207-210, 2009.

[21] H. Zhang, Y. Guo, and D. Zanotto, "Accurate ambulatory gait analysis in walking and running using machine learning models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 1, pp. 191-202, 2019.

[22] J. Figueiredo, C. P. Santos, and J. C. Moreno, "Automatic recognition of gait patterns in human motor disorders using machine learning: A review," *Medical engineering & physics,* vol. 53, pp. 1-12, 2018.

[23] P. Khera and N. Kumar, "Role of machine learning in gait analysis: a review," *Journal of Medical Engineering & Technology,* vol. 44, no. 8, pp. 441-467, 2020.

[24] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artificial Intelligence Review,* vol. 49, no. 1, pp. 1-40, 2018.

[25] S. Ilias, N. M. Tahir, R. Jailani, and C. Z. C. Hasan, "Classification of autism children gait patterns using neural network and support vector machine," in *2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2016: IEEE, pp. 52-56.

[26] K. Trentzsch *et al.*, "Using Machine Learning Algorithms for Identifying Gait Parameters Suitable to Evaluate Subtle Changes in Gait in People with Multiple Sclerosis," *Brain Sciences,* vol. 11, no. 8, p. 1049, 2021.

[27] S. Chakraborty, S. Jain, A. Nandy, and G. Venture, "Pathological gait detection based on multiple regression models using unobtrusive sensing technology," *Journal of Signal Processing Systems,* vol. 93, no. 1, pp. 1-10, 2021.

[28] G. Pendharkar, D. T. Lai, and R. K. Begg, "Detecting idiopathic toe-walking gait pattern from normal gait pattern using heel accelerometry data and support vector machines," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008: IEEE, pp. 4920-4923.

[29] G. Pendharkar, P. Percival, D. Morgan, and D. Lai, "Automated method to distinguish toe walking strides from normal strides in the gait of idiopathic toe walking children from heel accelerometry data," *Gait & posture,* vol. 35, no. 3, pp. 478-482, 2012.

[30] S. Kim, R. Soangra, M. Grant-Beuttler, and A. Aminian, "Wearable sensor-based gait classification in idiopathic toe walking adolescents," *Biomedical sciences instrumentation,* vol. 55, no. 2, p. 178, 2019.

[31] H. B. Menz, A. B. Dufour, J. L. Riskowski, H. J. Hillstrom, and M. T. Hannan, "Foot posture, foot function and low back pain: the Framingham Foot Study," *Rheumatology,* vol. 52, no. 12, pp. 2275-2282, 2013, doi: 10.1093/rheumatology/ket298.

[32] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE Trans Biomed Eng,* vol. 44, no. 3, pp. 136-47, Mar 1997, doi: 10.1109/10.554760.

[33] K. E. Atkinson, *An introduction to numerical analysis*. John wiley & sons, 2008.

[34] S. Yang, J.-T. Zhang, A. C. Novak, B. Brouwer, and Q. Li, "Estimation of spatio-temporal parameters for post-stroke hemiparetic gait using inertial sensors," *Gait & posture,* vol. 37, no. 3, pp. 354-358, 2013.

[35] G. Prateek, P. Mazzoni, G. M. Earhart, and A. Nehorai, "Gait cycle validation and segmentation using inertial sensors," *IEEE Transactions on Biomedical Engineering,* vol. 67, no. 8, pp. 2132-2144, 2019.

[36] J. Zhang, T. E. Lockhart, and R. Soangra, "Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors," *Ann Biomed Eng,* vol. 42, no. 3, pp. 600-12, Mar 2014, doi: 10.1007/s10439-013-0917-0.

[37] W. Si, G. Yang, X. Chen, and J. Jia, "Gait identification using fractal analysis and support vector machine," *Soft Computing,* vol. 23, no. 19, pp. 9287-9297, 2019.

[38] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST),* vol. 2, no. 3, pp. 1-27, 2011.

[39] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics,* vol. 21, no. 3, pp. 660-674, 1991.

[40] Y. Qi, "Random forest for bioinformatics," in *Ensemble machine learning*: Springer, 2012, pp. 307-323.

[41] L. E. Peterson, "K-nearest neighbor," *Scholarpedia,* vol. 4, no. 2, p. 1883, 2009.

[42] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer school on machine learning*, 2003: Springer, pp. 63-71.

[43] C. D. Brown and H. T. Davis, "Receiver operating characteristics curves and related decision measures: A tutorial," *Chemometrics and Intelligent Laboratory Systems,* vol. 80, no. 1, pp. 24-38, 2006.

[44] G. Pendharkar, P. Percival, D. Morgan, and D. Lai, "Automated method to distinguish toe walking strides from normal strides in the gait of idiopathic toe walking children from heel accelerometry data," *Gait Posture,* vol. 35, no. 3, pp. 478-82, Mar 2012, doi: 10.1016/j.gaitpost.2011.11.011.

[45] C. Christensen, A. Haddad, and E. Maus, "Validity of an Accelerometer Used to Measure Step Count in Children With Idiopathic Toe Walking," *Pediatr Phys Ther,* vol. 29, no. 2, pp. 153-157, Apr 2017, doi: 10.1097/PEP.0000000000000364.

[46] G. Pendharkar, P. Percival, and D. Morgan, "Evaluating bouncy gait in idiopathic toe-walkers using accelerometer," presented at the 2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2008.

[47] A. Kalron, L. Frid, S. Menascu, and U. Givon, "The association between gait variability with the energy cost of walking depends on the fall status in people with multiple sclerosis without mobility aids," *Gait & Posture,* vol. 74, pp. 231-235, 2019, doi: 10.1016/j.gaitpost.2019.09.021.

[48] J. D. Wong, J. C. Selinger, and J. M. Donelan, "Is natural variability in gait sufficient to initiate spontaneous energy optimization in human walking?," *Journal of Neurophysiology,* vol. 121, no. 5, pp. 1848-1855, 2019, doi: 10.1152/jn.00417.2018.

[49] D. B. Kowalsky, J. R. Rebula, L. V. Ojeda, P. G. Adamczyk, and A. D. Kuo, "Human walking in the real world: Interactions between terrain type, gait parameters, and energy expenditure," *PLoS One,* vol. 16, no. 1, p. e0228682, 2021, doi: 10.1371/journal.pone.0228682.

[50] P. G. Weyand, B. R. Smith, M. R. Puyau, and N. F. Butte, "The mass-specific energy cost of human walking is set by stature," *J Exp Biol,* vol. 213, no. Pt 23, pp. 3972-9, Dec 1 2010, doi: 10.1242/jeb.048199.

[51] C. B. Cunningham, N. Schilling, C. Anders, and D. R. Carrier, "The influence of foot posture on the cost of transport in humans," *J Exp Biol,* vol. 213, no. 5, pp. 790-7, Mar 1 2010, doi: 10.1242/jeb.038984.
[52] M. M. Eken, S. M. Braendvik, E. M. Bardal, H. Houdijk, A. J. Dallmeijer, and K. Roeleveld, "Lower limb muscle fatigue during walking in children with cerebral palsy," *Developmental Medicine & Child Neurology,* vol. 61, no. 2, pp. 212-218, 2019, doi: 10.1111/dmcn.14002.
[53] C.-W. Lung, B.-Y. Liau, J. A. Peters, L. He, R. Townsend, and Y.-K. Jan, "Effects of various walking intensities on leg muscle fatigue and plantar pressure distributions," *BMC Musculoskeletal Disorders,* vol. 22, no. 1, 2021, doi: 10.1186/s12891-021-04705-8.
[54] S. M. H. Tsang, G. P. Y. Szeto, L. M. K. Li, D. C. M. Wong, M. M. P. Yip, and R. Y. W. Lee, "The effects of bending speed on the lumbo-pelvic kinematics and movement pattern during forward bending in people with and without low back pain," *BMC Musculoskelet Disord,* vol. 18, no. 1, p. 157, Apr 17 2017, doi: 10.1186/s12891-017-1515-3.

Rahul Soangra is an Assistant Professor in the Department of Physical Therapy at Crean College of Health and Behavioral Sciences. He is also an affiliate faculty for Fowler School of Engineering at Chapman University. His research focuses on Gait and Posture biomechanics of patients with musculoskeletal and neurological disorders. His research interests include utilization of virtual environments and wearable sensors for gait rehabilitation.

Yuxin Wen received a BS degree in medical informatics and engineering from Sichuan University, Chengdu, China, in 2011, an MS degree in biomedical engineering from Zhejiang University, Hangzhou, China, in 2014, and a PhD degree in electrical and computer engineering from the University of Texas at El Paso, El Paso, TX, USA, in 2020. She is currently an assistant professor in the Dale E. and Sarah Ann Fowler School of Engineering at Chapman University, Orange, CA, USA. Her research interests are focused on statistical modeling, machine learning, health monitoring for manufacturing and healthcare applications.

Marybeth Grant-Beuttler, PT, PhD, PCS is an Associate Professor at the Crean College of Health and Behavioral Sciences. She earned her BS degree in Physical Therapy (1987) from Northwestern University Medical School, Chicago, Illinois, her MS in Human Movement Sciences (1997) University of North Carolina, her PhD in Movement Science (2004) Drexel University, Philadelphia, Pennsylvania, and her PCS Pediatric Clinical Specialist (2006) from the American Physical Therapy Association. Dr. Grant-Beuttler's areas of advanced study are pediatrics, motor control, motor learning, and motor development.

Hualin Yang received a dual BS degree from Chapman University, Orange, USA in 2021, with BS in Data Analytics, and BS in Business Administration. She is currently a venture capitalist specializing in the investment of AI and blockchain technologies. At the same time, she is an entrepreneur, Co-founded a decentralized autonomous organization (DAO) with two other partners. Her research focuses on machine learning based health monitoring and diagnosis.