Published in final edited form as:

IEEE Sens J. 2020 November ; 20(21): 12859–12870. doi:10.1109/jsen.2020.3000772.

Determining Physical Activity Characteristics from Wristband Data for Use in Automated Insulin Delivery Systems

Mert Sevil, Mudassir Rashid, Zacharie Maloney, Iman Hajizadeh, Sediqeh Samadi, Mohammad Reza Askari, Nicole Hobbs, Rachel Brandt, Minsun Park, Laurie Quinn, Ali Cinar

Mert Sevil, Rachel Brandt, Nicole Hobbs and Zacharie Maloney are with the Department of Biomedical Engineering (BME); Mudassir Rashid, Mohammad Reza Askari, Iman Hajizadeh and Sedigeh Samadi are with the Department of Chemical and Biological Engineering (ChBE); Ali Cinar is with the Departments of ChBE and BME, Illinois Institute of Technology, Chicago, IL 60616; Minsun Park and Laurie Quinn are with the College of Nursing, University of Illinois at Chicago, IL, 60616.

Abstract

Algorithms that can determine the type of physical activity (PA) and quantify the intensity can allow precision medicine approaches, such as automated insulin delivery systems that modulate insulin administration in response to PA. In this work, data from a multi-sensor wristband is used to design classifiers to distinguish among five different physical states (PS) (resting, activities of daily living, running, biking, and resistance training), and to develop models to estimate the energy expenditure (EE) of the PA for diabetes therapy. The data collected are filtered, features are extracted from the reconciled signals, and the extracted features are used by machine learning algorithms, including deep-learning techniques, to obtain accurate PS classification and EE estimation. The various machine learning techniques have different success rates ranging from 75.7% to 94.8% in classifying the five different PS. The deep neural network model with long short-term memory has 94.8% classification accuracy. We achieved 0.5 MET (Metabolic Equivalent of Task) root-mean-square error for EE estimation accuracy, relative to indirect calorimetry with randomly selected testing data (10% of collected data). We also demonstrate a 5% improvement in PS classification accuracy and a 0.34 MET decrease in the mean absolute error when using multi-sensor approach relative to using only accelerometer data.

Graphical Abstract

Index Terms—

Wearable sensors; sensor fusion; physical activity classification; energy expenditure estimation; automated insulin delivery

I. Introduction

INCORPORATING physical activity (PA) in daily life provides several health benefits, such as reduced cardiovascular risk factors and improved well-being. PA also promotes metabolic health and glycemic control by increasing the effect of insulin on glucose uptake. Blood glucose responses to PA in people with Type 1 diabetes (T1D) are highly variable, and the imbalance in glucose availability and utilization during PA can increase the risk of dangerously low blood glucose levels (hypoglycemia) [1]–[3]. Therefore, people with T1D must make insulin dose adjustments to respond to the diverse modalities, intensities, and durations of PA [1]–[3]. Automated insulin dosing systems must also be informed of PA to adjust the computations of the proper insulin dose. PA increases the blood glucose consumption rate and excess insulin can further escalate blood glucose utilization, which can increase the risk of hypoglycemia during or several hours after PA in people with T1D [1]–[4]. Informing the insulin dose calculations of the PA characteristics is possible through the use of physiological data collected in real time from wearable devices [5]–[14]. The physiological variables measured by the sensors in wearable devices must be processed to obtain valuable information for the medical dosing calculations [10]–[14]. Knowledge of the type, intensity, and duration of physical activities can be obtained by detecting and interpreting the physical state (PS) and estimating the energy expenditure (EE), leading to the preventive and therapeutic benefits for people with chronic diseases [3], [15]–[24].

Many consumer products provide good enough accuracy for tracking various daily activities and ascertaining summary health statistics [10]–[14], [25]–[30]. Several studies monitor PA using mobile phones and sensors embedded in clothing, and some studies compare the effects of the wearable sensor location on the accuracy of PA tracking [31]–[38]. PA information in medical applications such as insulin dose decisions requires stringent levels of accuracy and continuous use in free-living conditions with real-time data streaming [1], [3], [39]–[46]. Several studies report EE estimation exclusively from accelerometer sensors [47], [48], and some commercial products also provide EE estimates [5], [49]. A few studies use chestbands, armbands, and experimental devices to estimate the EE [50]–[57]. Despite the proliferation of numerous devices, wrist-worn PA trackers are preferred in applications where requirements include the convenience of the device and the continuous use in free-living and diverse ambulatory conditions encountered in daily life [20]–[24], [58].

In this study, data from the Empatica E4 [6] wristband device comprising of several sensors (Table I) are used with a sensor fusion approach and machine learning (ML) techniques to determine the PS of the subject and estimate the EE during the ongoing PA. We demonstrate that employing several wristband sensors (skin conductance and temperature, accelerometers, and blood volume pulse) and integrating the sensor data can improve the accuracy of the PS classification and the EE estimation compared to the use of a single

sensor (for instance, accelerometers), resulting in more reliable PA information for therapeutic decision making [59]–[64]. Compared to the use of a single sensor (accelerometers), we show significant improvements in the PA classification (the accuracy increases by approximately 5%) and EE estimates (the mean absolute percentage deviation decreases by approximately 10%) when multiple sensors are integrated and used simultaneously for tracking PA. The PS considered in this work (sedentary state (SS), bicycling (BK), resistance training (RT), running (RN), and activities of daily life (DA)) are classified with an accuracy of 94.8%. The Empatica E4 device measures skin temperature (ST), 3-dimensional accelerometer (ACC), galvanic skin response (GSR), and blood volume pulse (BVP) based on photoplethysmography (PPG) from which heart rate (HR) is derived [6]. Features are extracted from these biosignals and the feature variables are used in ML algorithms to determine the PS and EE. The EE estimation accuracy of the proposed algorithms is compared to data from indirect calorimetry method using the COSMED Omnia K5 metabolic system [65], and the results demonstrate a low mean absolute error of 0.25 MET relative to the comparison indirect calorimetry method [65]. Over 430 h of experiments are conducted involving 25 subjects participating in the five PS considered for classification, and the accuracies of the various ML techniques are assessed.

II. Procedure

A. Data Collection

Wristband data are collected during five different PS (SS, BK, RT, RN, and DA). A total of 25 subjects (12 male/13 female) participated in a subset of the physical activities during the experiments. The subject demographic information (mean $\pm$ standard deviation) are: age 24.88 ± 3.15 yr, weight 72.42 ± 14.15 kg, height 1.70 ± 0.09 m, body mass index 25.00 ± 4.97 kg/m². Some subjects repeated certain physical activities, resulting in a number of experiments for each of the 25 subjects involved in the study (Table II). All experiment protocols were approved by the Institutional Review Board.

SS experiments are conducted while subjects are laying down or sitting and resting, and include activities like watching a movie, surfing the internet, reading a book, attending class lectures, working on a computer, and attending meetings. Seven subjects participated in 96 experiments for the SS data collection (Table II). The duration of the SS experiments is relatively longer than other PA because of the lack of fatigue and protracted comfort of the subjects during resting conditions relative to the other PS. Data for DA are collected from 7 subjects in 43 experiments including walking in the city, walking on the treadmill (1.7 mph – 3.5 mph), washing dishes, house cleaning, teaching a class, and shopping [66]–[71]. RN experiments are conducted mostly on the treadmill and includes a few outdoor runs. RN data include 180 experiments conducted by 21 different subjects with different speeds in the range of 3.5 – 8.0 mph [66]–[71]. The thresholds for walking and running were selected to correspond with the criteria used in existing literature studies [66]–[71]. BK experiments are primarily conducted with a stationary bike except for a few outdoors biking activities, with a total of 56 BK experiments conducted involving 16 different subjects at variable intensities in the range of 50 – 150 W.

RT activities include weightlifting, dumbbell chest press, leg extension, leg curl, lat pull down, dumbbell shoulder press, and seated row (rowing machine workout). Approximately 17 h of RT data are collected from 8 subjects. All data are recorded with labeled PS information. Overall, over 430 h of data from 25 subjects are collected for common types of daily PA from the large subject group [3], [39], [72]–[79].

In addition to the wristband data (Empatica E4) [6], indirect calorimetry data are collected with the COSMED Omnia K5 metabolic system during the 98 EE data collection experiments (Table II) to develop the EE estimation model [65]. The duration of data collected with the K5 metabolic system and E4 wristband is shorter than data collected using the E4 wristband only due to concerns for comfort of the subjects. Overall 80 h of experiments are conducted with the K5 metabolic system.

B. Pre-Processing

The Empatica E4 is a research grade device used in several studies to monitor subjects throughout daily life for events necessitating medical decisions or interventions [80]–[89]. The ACC, BVP, GSR, ST, and HR biosignals are measured by the Empatica E4 at 64, 32, 4, 4 and 1 samples per second, respectively [6]. Although more frequent sampling and raw data collection of the measurements can provide better tracking of PA in shorter time scales, the objective of this work is to use the PA information in insulin dosing decisions made every one to five minutes. Therefore, we use the Empatica E4 measurements and integrate them to determine the average PS and EE over each one-minute interval.

Different filters are used to eliminate noise from the collected data. A median filter and a Savitzky-Golay filter are chosen for ACC, HR, ST, and GSR data filtering [61]. The BVP signal is corrupted by motion artifacts and environmental noise, thus a cascaded signal processing technique is used, including band-pass filter (Butterworth), nonlinear recursive least squares for adaptive noise cancellation, and wavelet decomposition [90], [91]. The use of the sequential signal processing techniques reduced the correlation between the raw BVP signal and the denoised BVP signal from 0.258 to 0.008, demonstrating that the cleaned and processed BVP data are uncorrelated with the accelerometer measurements (Fig. 1).

C. Feature Extraction

The filtered biosignals are used to extract features at each one-minute interval. A comprehensive literature review is conducted to determine the appropriate feature extraction methods, including statistical parameters (standard deviation, median, max, kurtosis, skewness, etc.), mathematical functions (mean, differentiation, arccosine, etc.) and data-specific calculations (zero-cross of ACC readings, total-energy response of GSR, maximum amplitude of the very low frequency response of BVP, etc.) [61], [92]–[99]. Furthermore, we generated new features as the ratio of extracted features, which can be informative to distinguish the PS. For example, during BK the ACC readings (average magnitude of ACC) are relatively stable, yet the GSR readings (average magnitude of GSR) increase because of the higher sweating rate. Therefore, the ratio of the average magnitudes of ACC to GSR can be informative for discriminating the BK exercise from the other PS. To our knowledge, this has not been reported previously. The features pool includes 866 primary features extracted

from 3D ACC, raw BVP, processed denoised BVP [90], HR, ST, and GSR readings, with 225, 148, 148, 98, 111, and 136 primary features, respectively. The same features are extracted from both raw and processed BVP readings because the raw BVP signal includes motion artifacts that can aid in discriminating certain PS [90]. In addition, the preprocessed signal represents HR variability independent of the motion artifacts, thus both the raw and preprocessed HR signals can provide informative features. An additional 1,350 features are obtained from the ratios of the primary features, bringing the total features to 2,216. A list of selected features that are informative for PS classification and EE estimation is provided in Table III.

During the SS experiments, a relatively larger dataset is collected. To avoid the negative impact of imbalances in the class sizes, the feature variables data for the other PS are upsampled by generating synthetic samples using the adaptive synthetic sampling approach (ADASYN) [100]. Finally, 102,950 min of data are obtained; 16,758, 20,747, 23,886, 18,712, and 22,847 mins of data for the SS, DA, RT, BK, and RN states, respectively.

Physiological time delay exists for some of the measurements during exercise, such as the delayed increase of the GSR relative to the onset of PA (Fig. 2). We consider the inherent time delay in the measurements within the modeling paradigms by incorporating time-delayed versions of the extracted features in the dataset, which enables consideration of timed delays of up to 15 min for the measurements. Therefore, any time delay of up to 15 min in the physiological measurements is captured by the models.

The data collected are divided into three mutually exclusive sets for training (80% of data), validating (10% of data), and testing (10% of data) our models and algorithms.

D. Dimension Reduction

The same pool of feature variables is used for PS classification and EE estimation. Most of the extracted features are highly correlated with each other (such as median of HR and mean of HR). Highly correlated features do not provide additional information for ML algorithms, and their use may cause bias to the same attributes, leading to poor estimation and classification accuracy. Principal component analysis (PCA) is used to reduce the dimensionality of the extracted features [101]. We determine all possible combinations of PS classes and compute the p-values for each feature variable between the PS classes. Therefore, we evaluate whether the feature variables are statistically different across pairs of classes, for all possible combinations of PS. The feature variables are retained for model development when the p-values from a two-sided t-test depict a significant difference across PS classes at a significance level of $\alpha = 1\%$. We use the more stringent significance level criteria relative to $\alpha = 5\%$ to ensure a sufficient difference in the feature variables across the PS. Selected features (1730 features) are used in the PCA dimension reduction [102]. The number of retained principal components (PC) is optimized by varying the variance explained from 80 to 99% of the total variance in the feature variable space. Based on the classification accuracy, the first 175 PCs are retained, capturing over 95% of the variation in the overall features. The retained PCs are then normalized to range from 0 to 1 [103].

Partial least squares (PLS) regression is used to reduce the features to a smaller set of latent variables (LV) that are correlated with the EE [101]. The variance explained by the retained LVs is varied from 80 to 99%, and the number of LVs to be retained are determined based on the cross-validation of EE estimation accuracy. The first 75 LVs are retained, explaining over 97% of the total variation in all the extracted feature variables. The latent variables are normalized to range between 0 and 1.

E. Machine Learning Algorithms

The normalized PCs/LVs (inputs) and the corresponding numerically labeled PS information recorded (outputs) are used with supervised ML algorithms to classify the PS. The ML algorithms used include k-nearest neighbors (k-NN), support vector machines (SVM), naive Bayes (NB), decision tree (DT), neural networks (NN), linear discrimination (LD), ensemble learning (EL), and deep learning with long short-term memory (LSTM-DL). The models of the normalized LVs (inputs) and the corresponding gold-standard (from K5 metabolic system data) EE values reported in MET (Metabolic Equivalent of Task) (outputs) are trained and validated with several regression models including support vector regression (SVR), k-NN, linear regression (LR), DT, NN, Gaussian process regression (GPR), EL and LSTM-DL.

1) k-nearest neighbors: We used hyperparameter optimization to determine distance [104], which is found as the cosine distance metric to compare relative distances between the feature variables of the testing data with the feature variables of the training data. Based on optimization results, the 10 neighbors with the shortest sorted distances are identified as being representative of each class. We assign the class label to the new data according to the most common class within 10 closest training samples.

2) Support Vector Machine/Regression: The SVM have the advantage of easy adaptation to a feature-based classification approach for feature variables with nonlinear relations [105]. SVM functions are defined by determining a separating hyperplane in the high-dimensional feature space that best distinguishes two classes at the training stage. The separating hyperplane determined is used on-line with testing data to assign a class to the test data sample, based on the feature variable inputs [105]. We used SVM/SVR both to determine PS and estimate EE [105].

3) Decision Tree: DT generates a predictive model that can be used as classification or regression model [106]. DT algorithms are sensitive to variations in data and relatively unstable [106].

4) Naive Bayes: NB models are based on statistical information and compute the variance and mean of the selected subset of feature variables for a specific number of clusters [107]. We compute the Gaussian and kernel probability density for each feature, and we compute posterior probability values for the clusters [107]. The computed posterior values are compared to determine the classification.

5) Neural Network: NN can capture nonlinear relations with multiple hidden layers [108]. Overfitting and predicting good initial parameters (number of layers, number of neurons, activation function) are the major challenges, which can be addressed by optimization. We used ordinary two-layer NN (15 hidden neurons in each layer) by utilizing the selection criteria (Eq. (1)) [109] for both PS classification and EE estimation

$$NHN = \frac{Ns}{\gamma(Ni + No)} \quad (1)$$

where N_i is the number of inputs, N_o is the number of outputs, N_s is the number of samples, and γ is an arbitrary scaling factor usually in the range 2–10 [109].

6) Gaussian Process Regression: GPR is a stochastic statistical technique that seeks a multivariate normal distribution within random variables from the feature space domain. GPR measures the similarity between points (the kernel function) to predict the value from training data [110]. We used GPR to estimate EE.

7) Ensemble Learning: EL uses trained multiple learners, in this case multiple decision trees, to achieve better accuracy than the individual models can achieve alone. Different methods are used to establish an ensemble learning model including boosting and bagging method [111]. Hyperparameters optimization helps to select the best method with the optimum number of learning cycle, learning rate and minimum leaf size. We developed an ensemble learning model to both classify PS and estimate EE [111].

8) Linear Discrimination/Regression: LR uses features to create a linear model [112]. LD finds a linear combination of features that characterizes or separates two or more classes of objects or events. It is unable to capture nonlinearity. Physiological responses may contain some nonlinearities that cannot be handled with a linear model. We developed linear models to compare their classification and estimation performance with the other ML techniques [112].

9) Long-Short Term Memory and Deep Learning: Deep learning (DL) is a combination of different layers including fully connected, long short-term memory (LSTM), softmax, regression, Rectified linear unit (ReLU), and dropout layers [113]. We evaluated various network structures to determine the most accurate deep NN for PS classification and EE estimation. Fig. 3 illustrates our proposed DL structure for PS classification and EE estimation. In addition, we used the L2 regularization term to reduce the risk of overfitting (value: 0.05).

LSTM is a recurrent NN (RNN) architecture used in DL [114], [115]. Long-term dependencies are a significant problem which could not be solved with ordinary RNN. A modification of the ordinary RNN structure enables handling this problem with the solution of vanishing gradient problem. It commonly contains input gate (i), cell (C_t), update gate, forget gate (f), output gate (o), hidden state (h_t) and memory cell (g) (Fig. 4) [114], [115]. Update gate is a combination of the output of f, i, and g. LSTM enables evaluation of past and current information without long-term dependencies. In addition to current information,

previously extracted features of biosignals have been considered in the LSTM to capture physiological alterations more rapidly. Deeper LSTM (two LSTM layers) is used with 30 hidden neurons for classification. We used a single LSTM layer with 10 hidden neurons for EE estimation. The numbers of hidden neurons (NHN) are selected based on Eq. (1).

Dropout layer drops out hidden and visible units of network randomly, commonly used for regularization to reduce the risk of overfitting [113]. The softmax function can be used as a logistic regression model for the multi-classification task [113]. It transforms the numerical probability outputs of the network into the discrete class labels. It is commonly used as a last layer of a DNN. The fully connected layer is an ordinary hidden layer, which is generally used after LSTM layer. It has five hidden neurons (equal to the number of PS classified) for PS classification, and a single hidden neuron for EE estimation. ReLU layer performs a threshold operation to each element of the input, where any value less than zero is set to zero. It is commonly used as an activation function to improve the learning ability of the network [113]. Regression layer is commonly the last layer of estimation-based deep NN designs. It converts all predicted probabilities to actual expected output estimation value [113].

10) Hyperparameter Optimization: The various hyperparameters for the different ML techniques are optimized with Bayesian optimization (expected improvement acquisition function) to achieve the best classification accuracy [116]. Table IV lists the hyperparameters for each ML algorithm. A large number iterations (100 iterations) are performed for each ML algorithm using 4-fold cross-validation techniques.

III. Results

The results are evaluated for the PS classification and the EE estimation algorithms using the testing data set and by employing various ML metrics.

A. PS Classification

We evaluated different ML algorithms (Fig. 5). LSTM-DL achieved a 94.8% accuracy with the test dataset, which was better than the other algorithms as the next-best algorithm (EL) achieved a classification accuracy of 93.2%. LSTM has the ability of bridging long time-lags between inputs. The LD and NB algorithms have lower classification performances below 80% with the same dataset and inputs.

Precision and F-score are important common ML metrics. Table V reports the values of the metrics for different ML algorithms used for all PS. The receiving operating characteristic (ROC) curve provides a good illustration of recall, precision, and F-score with a single plot presenting the performances of the classification models (Fig. 6).

We also compare the improvement in PS classification accuracy obtained by the proposed sensor fusion approach with an single sensor approach employing only ACC data. We generate features from the ACC data and train models with the features to generate algorithms relying exclusively on ACC data. Using ACC data only results in an accuracy of

approximately 90%, which is lower than the almost 95% accuracy obtained through the sensor fusion approach integrating the five different biosignal simultaneously (Table VI).

B. EE Estimation

The root-mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) are calculated for the estimated EE compared to the indirect calorimeter measurements for the different ML algorithms (Table VII) for the testing dataset (10%). LSTM-DL performed better than the other algorithms with an MAE of 0.25 MET compared to 0.26 MET for the GPR, which has the best accuracy among the other algorithms.

We also evaluate the accuracy of the EE estimated when using the proposed sensor fusion approach against the case where a single sensor measurement is used, in this case only ACC data. We again generate features from the ACC data only for comparison, and evaluate it against the situation where features from all measured variables are integrated for EE estimation. The estimation error in terms of RMSE increased from 0.51 to 0.79 MET when only ACC data is used, or in MAPE from 12.1 to 22.0%, and in MAE from 0.25 to 0.59 MET (Table VI). Therefore, the use of multiple sensors and a sensor fusion approach results in better EE estimation accuracy in comparison to a single sensor approach.

IV. Discussion of the Results

A. PS Classification

The confusion matrix for LSTM-DL (Fig. 7) illustrates all misclassifications for each PS. The SS class is the most accurately classified PS because accelerometer readings are mostly stable and SS is easily distinguished from other states even when relying solely on the ACC data. A few BK activities are misclassified as SS (0.2% of the BK class samples are misclassified, a total of 21 mins out of 2225 mins) because during both BK and SS classes, the ACC readings are relatively stable (holding the handlebar of the bike). The SS activities are the most misclassified state (87.5% accuracy) because some low-intensity DA have similar biosignals readings and characteristics as the SS class. The GSR and ACC measurements have higher standard deviations during RT relative to other activities because the measurement values for GSR and ACC decrease substantially during the resting period between sets of repetitions. For example, the standard deviation of GSR across a RT experiment was found to be $0.99 \mu S$ compared to a maximum of $0.60 \mu S$ for other activities. The variability in the standard deviation of GSR and ACC is not the only indicator of RT class. The mean of ACC and GSR readings are also relatively lower during RT compared to RN or higher-intensity DA. For example, the mean of ACC and GSR across RT experiments were found to be 65.5 g and $0.79 \mu S$ compared to an average of 63.9 g and $0.06 \mu S$ for SS activities, respectively.

Overall, the proposed method achieved 94.8% accuracy (Fig. 7) with the data collected from 25 participants and 430 h of experiments with five different PS. The results compare favorably against similar recent studies, with reported accuracies ranging between 95 to 98%

using sensors in smart garments or shoes, and with novel devices integrating accelerometers, gyroscopes, and barometric pressure sensors [3], [35]–[38], [55], [67], [117]–[121].

A limitation of the proposed work is the number of subjects involved in the study and the classification of only five PA. The accuracy of the proposed methods can be improved by collecting more data from numerous subjects. The discrepancies in the number of subjects due to the PA preferences of individuals can lead to unbalanced class sizes and biased results, which we tried to mitigate by using upsampling techniques. Future work will augment the data with new subjects, particularly for exercises where few subjects performed the activity.

B. EE Estimation

Independent of the training, testing, and validation experiments, a new experiment is conducted for circuit training of approximately 110 min to evaluate the EE estimation accuracy. The experiment includes stationary bike (20 min), resting (10 min), rowing machine workout (20 min), resting (10 min), treadmill running (20 min), resting (10 min) and activities of daily living (20 min). The treadmill speed fluctuates between 3.5 and 6.0 mph at 0% incline. Stationary bike intensity is in the range of 50 – 110 W. Activities of daily living are emulated with treadmill walking at low speed in the range of 1.5–3.5 mph with 0% incline. Resistance exercise is done with a rowing machine workout in the range of 50 – 75 W. During resting periods, the subject sat in a chair and avoided any physical activities. Indirect calorimeter measurements and estimated EE values (LSTM-DL) shown in Fig. 8 illustrate that the EE estimation algorithms can track EE accurately over the range of physical activities.

Overall, the EE is estimated with approximately 0.5 MET RMSE over the 8 h testing dataset including different types of medium-intensity exercises. A separate experiment (approximately 2 h (110 min)) is conducted to verify our model, which yielded similar findings as the testing dataset (Fig. 8). Our method can achieve competitive results relative to values reported in recent literature focusing on EE estimation (Table VI) [50]–[57], [122]. The MAE of 0.25 MET achieved by our method is competitive with the literature reported values. Moreover, our results are based on a convenient and practical wristband physical activity tracker, thus suitable for use in free-living ambulatory settings.

C. Extensions of Proposed Method

Our method is based on a convenient and practical wristband device, similar to popular consumer wearable wristband activity trackers [10]–[14], [25]–[28]. This makes it an appealing solution for monitoring the PA type and intensity in medical applications [3], [20]–[24]. The high accuracy of the results from a wristband device means the approach can be useful for clinical decision making in managing chronic diseases such as diabetes and obesity [15]–[19]. The PS and EE information generated from the refined and processed signals can be used to inform automated insulin delivery systems to automatically modulate the insulin dosing based on the type and intensity of the PA [1], [3], [39]–[46]. Specifically, if an increase in the EE is detected, then the insulin doses can be decreased or suspended to avoid excess insulin delivery that may lead to an adverse low glucose values. Predictions of

future glucose trajectory profiles are frequently used in managing Type 1 diabetes, and the models used for making the glucose predictions can be augmented with additional supplementary inputs like EE values to increase the accuracy of the glucose forecasts. The automated insulin dosing systems rely on control algorithms to determine the insulin dose amount, and using the PS and EE estimates to modify the control parameters by making the control algorithm suggest less insulin when EE is elevated can be instrumental in improving the lives of people using automated insulin delivery systems.

V. Conclusion

An approach to detect the PS and estimate the EE through physiological signals reported by a single wristband device with multiple sensors is presented. The proposed method achieved a 95% accuracy for PS classification and a RMSE of 0.51 MET for EE estimation. We also demonstrated that there is a significant improvement in PS classification and EE estimation using multiple sensors simultaneously in a sensor fusion approach as opposed to the use of a single sensor for monitoring PA. The accuracy of the PS classification and EE estimation results by using the sensors in a convenient wristband activity tracker will enable precision medicine by considering physical activity type and intensity in computing insulin dosing decisions.

Acknowledgments

Financial support from the NIH under the grants 1DP3DK101075 and 1DP3DK101077 and JDRF under grant 2-SRA-2017-506-M-B made possible through collaboration between the JDRF and The Leona M. and Harry B. Helmsley Charitable Trust is gratefully acknowledged.

References

- [1]. Riddell MC, Zaharieva DP, Yavelberg L, Cinar A, and Jamnik VK, "Exercise and the development of the artificial pancreas: one of the more difficult series of hurdles," *Journal of Diabetes Science and Technology*, vol. 9, no. 6, pp. 1217–1226, 2015. [PubMed: 26428933]
- [2]. Riddell MC, Gallen IW, Smart CE, Taplin CE, Adolfsson P, Lumb AN, Kowalski A, Rabasa-Lhoret R, McCrimmon RJ, Hume C et al., "Exercise management in type 1 diabetes: a consensus statement," *The Lancet Diabetes & Endocrinology*, vol. 5, no. 5, pp. 377–390, 2017. [PubMed: 28126459]
- [3]. Turksoy K, Paulino TML, Zaharieva DP, Yavelberg L, Jamnik V, Riddell MC, and Cinar A, "Classification of physical activity: information to artificial pancreas control systems in real time," *Journal of Diabetes Science and Technology*, vol. 9, no. 6, pp. 1200–1207, 2015. [PubMed: 26443291]
- [4]. Colberg SR, *Exercise and Diabetes: A Clinician's Guide to Prescribing Physical Activity*. American Diabetes Association, 2013.
- [5]. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, Schwartz JE, and Davidson KW, "Fitbit: An accurate and reliable device for wireless physical activity tracking," *International Journal of Cardiology*, vol. 185, pp. 138–140, 2015. [PubMed: 25795203]
- [6]. McCarthy C, Pradhan N, Redpath C, and Adler A, "Validation of the empatica e4 wristband," in *Engineering in Medicine and Biology Society. IEEE*, 2016, pp. 1–4.
- [7]. Patel S, Park H, Bonato P, Chan L, and Rodgers M, "A review of wearable sensors and systems with application in rehabilitation," *Journal of Neuroengineering and Rehabilitation*, vol. 9, no. 1, p. 21, 2012. [PubMed: 22520559]
- [8]. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, Damico R, Slomian J, Reginster JY, Bruyere O et al., "Smart wearable body sensors for patient self-

assessment and monitoring,” *Archives of Public Health*, vol. 72, no. 1, p. 28, 2014. [PubMed: 25232478]

- [9]. Dobkin BH, “Wearable motion sensors to continuously measure real-world physical activities,” *Current Opinion in Neurology*, vol. 26, no. 6, p. 602, 2013. [PubMed: 24136126]
- [10]. Esfahani MIM and Nussbaum MA, “A smart undershirt for tracking upper body motions: task classification and angle estimation,” *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7650–7658, 2018.
- [11]. Mukhopadhyay SC, “Wearable sensors for human activity monitoring: A review,” *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1321–1330, 2014.
- [12]. Incel OD, Kose M, and Ersoy C, “A review and taxonomy of activity recognition on mobile phones,” *BioNanoScience*, vol. 3, no. 2, pp. 145–171, 2013.
- [13]. Shoaib M, Bosch S, Incel O, Scholten H, and Havinga P, “A survey of online activity recognition using mobile phones,” *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015. [PubMed: 25608213]
- [14]. Schrack JA, Cooper R, Koster A, Shiroma EJ, Murabito JM, Rejeski WJ, Ferrucci L, and Harris TB, “Assessing daily physical activity in older adults: unraveling the complexity of monitors, measures, and methods,” *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 71, no. 8, pp. 1039–1048, 2016.
- [15]. Millar WJ and Young TK, “Tracking diabetes,” *Public Health Reports*, vol. 14, no. 3, p. 1, 2003.
- [16]. Morey MC, Pieper CF, Edelman DE, Yancy WS Jr, Green JB, Lum H, Peterson MJ, Sloane R, Cowper PA, Bosworth HB et al., “Enhanced fitness: a randomized controlled trial of the effects of home-based physical activity counseling on glycemic control in older adults with prediabetes mellitus,” *Journal of the American Geriatrics Society*, vol. 60, no. 9, pp. 1655–1662, 2012. [PubMed: 22985140]
- [17]. Yang X, Telama R, Viikari J, and Raitakari OT, “Risk of obesity in relation to physical activity tracking from youth to adulthood,” *Medicine and Science in Sports and Exercise*, vol. 38, no. 5, p. 919, 2006. [PubMed: 16672846]
- [18]. Twisk J, Kemper H et al., “Tracking of activity and fitness and the relationship with cardiovascular disease risk factors,” *Medicine and Science in Sports and Exercise*, vol. 32, no. 8, pp. 1455–1461, 2000. [PubMed: 10949012]
- [19]. Hills AP, Andersen LB, and Byrne NM, “Physical activity and obesity in children,” *British Journal of Sports Medicine*, vol. 45, no. 11, pp. 866–870, 2011. [PubMed: 21836171]
- [20]. Dominick GM, Winfree KN, Pohlig RT, and Papas MA, “Physical activity assessment between consumer-and research-grade accelerometers: a comparative study in free-living conditions,” *JMIR mHealth and uHealth*, vol. 4, no. 3, p. e110, 2016. [PubMed: 27644334]
- [21]. Brooke SM, An H-S, Kang S-K, Noble JM, Berg KE, and Lee J-M, “Concurrent validity of wearable activity trackers under free-living conditions,” *Journal of strength and conditioning research*, vol. 31, no. 4, pp. 1097–1106, 2017. [PubMed: 27465631]
- [22]. Campos-Meirinhos EJ, Mayorga-Vega D, Casado-Robles C, Guijarro-Romero S, and Viciania J, “Are activity wristbands valid to estimate moderate-to-vigorous physical activity in adolescents during free-living conditions?” *Motricidade*, vol. 15, pp. 18–18, 2019.
- [23]. Bassett DR, “Device-based monitoring in physical activity and public health research,” *Physiological Measurement*, vol. 33, no. 11, p. 1769, 2012. [PubMed: 23110847]
- [24]. Bonomi A and Westerterp K, “Advances in physical activity monitoring and lifestyle interventions in obesity: a review,” *International Journal of Obesity - Nature*, vol. 36, no. 2, p. 167, 2012.
- [25]. Adam Noah J, Spierer DK, Gu J, and Bronner S, “Comparison of steps and energy expenditure assessment in adults of fitbit tracker and ultra to the actual and indirect calorimetry,” *Journal of Medical Engineering and Technology*, vol. 37, no. 7, pp. 456–462, 2013. [PubMed: 24007317]
- [26]. Bunn JA, Navalta JW, Fountaine CJ, and Reece JD, “Current state of commercial wearable technology in physical activity monitoring 2015–2017,” *International Journal of Exercise Science*, vol. 11, no. 7, p. 503, 2018. [PubMed: 29541338]
- [27]. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, and Lynch C, “Validation of the fitbit one, garmin vivofit and jawbone up activity tracker in estimation of energy expenditure during treadmill walking and running,” *Journal of Medical Engineering and Technology*, vol. 41, no. 3, pp. 208–215, 2017. [PubMed: 27919170]

- [28]. Soutiere SE, Cox BD, Laird MD, Markwald RR, Heaney JH, Chinoy ED, and Simmons RG, "Wearable activity tracker literature review (2009–2016)," Naval Health Research Center, Tech. Rep, 2017.
- [29]. Hedegaard M, Anvari-Moghaddam A, Jensen BK, Jensen CB, Pedersen MK, and Samani A, "Prediction of energy expenditure during activities of daily living by a wearable set of inertial sensors," *Medical Engineering and Physics*, vol. 75, pp. 13–22, 2020. [PubMed: 31679905]
- [30]. Shwetar YJ, Veerubhotla AL, Huang Z, and Ding D, "Comparative validity of energy expenditure prediction algorithms using wearable devices for people with spinal cord injury," *Spinal Cord - Nature*, pp. 1–10, 2020.
- [31]. Cleland I, Kikhia B, Nugent C, Boytsov A, Hallberg J, Synnes K, McClean S, and Finlay D, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors*, vol. 13, no. 7, pp. 9183–9200, 2013. [PubMed: 23867744]
- [32]. Mokhlespour Esfahani M and Nussbaum M, "Preferred placement and usability of a smart textile system vs. inertial measurement units for activity monitoring," *Sensors*, vol. 18, no. 8, p. 2501, 2018.
- [33]. Boerema S, van Velsen L, Schaake L, Tönis T, and Hermens H, "Optimal sensor placement for measuring physical activity with a 3d accelerometer," *Sensors*, vol. 14, no. 2, pp. 3188–3206, 2014. [PubMed: 24553085]
- [34]. Özdemir A, "An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice," *Sensors*, vol. 16, no. 8, p. 1161, 2016.
- [35]. Ahmadi MN, Pfeiffer KA, and Trost SG, "Physical activity classification in youth using raw accelerometer data from the hip," *Measurement in Physical Education and Exercise Science*, pp. 1–7, 2020.
- [36]. Manjarres J, Narvaez P, Gasser K, Percybrooks W, and Pardo M, "Physical workload tracking using human activity recognition with wearable devices," *Sensors*, vol. 20, no. 1, p. 39, 2020.
- [37]. Sheng B, Moosman OM, Del Pozo-Cruz B, Del Pozo-Cruz J, Alfonso-Rosa RM, and Zhang Y, "A comparison of different machine learning algorithms, types and placements of activity monitors for physical activity classification," *Measurement*, p. 1, 2020.
- [38]. Sedighi H, "Classification of human activity recognition using smart-phones," *ArXiv*, p. 1, 2020.
- [39]. Turksoy K, Quinn L, Littlejohn E, and Cinar A, "Multivariable adaptive identification and control for artificial pancreas systems," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 883–891, 2013.
- [40]. Turksoy K, Hajizadeh I, Samadi S, Feng J, Sevil M, Park M, Quinn L, Littlejohn E, and Cinar A, "Real-time insulin bolusing for unannounced meals with artificial pancreas," *Control Engineering Practice*, vol. 59, pp. 159–164, 2017.
- [41]. Turksoy K, Monforti C, Park M, Griffith G, Quinn L, and Cinar A, "Use of wearable sensors and biometric variables in an artificial pancreas system," *Sensors*, vol. 17, no. 3, p. 532, 2017.
- [42]. Breton MD, Brown SA, Karvetski CH, Kollar L, Topchyan KA, Anderson SM, and Kovatchev BP, "Adding heart rate signal to a control-to-range artificial pancreas system improves the protection against hypoglycemia during exercise in type 1 diabetes," *Diabetes Technology and Therapeutics*, vol. 16, no. 8, pp. 506–511, 2014. [PubMed: 24702135]
- [43]. Turksoy K, Quinn LT, Littlejohn E, and Cinar A, "An integrated multivariable artificial pancreas control system," *Journal of Diabetes Science and Technology*, vol. 8, no. 3, pp. 498–507, 2014. [PubMed: 24876613]
- [44]. Ding S and Schumacher M, "Sensor monitoring of physical activity to improve glucose management in diabetic patients: a review," *Sensors*, vol. 16, no. 4, p. 589, 2016.
- [45]. Jacobs PG, Resalat N, El Youssef J, Reddy R, Branigan D, Preiser N, Condon J, and Castle J, "Incorporating an exercise detection, grading, and hormone dosing algorithm into the artificial pancreas using accelerometry and heart rate," *Journal of Diabetes Science and Technology*, vol. 9, no. 6, pp. 1175–1184, 2015. [PubMed: 26438720]
- [46]. Hajizadeh I, Rashid M, Turksoy K, Samadi S, Feng J, Sevil M, Hobbs N, Lazaro C, Maloney Z, Littlejohn E et al., "Incorporating unannounced meals and exercise in adaptive learning of personalized models for multivariable artificial pancreas systems," *Journal of Diabetes Science and Technology*, vol. 12, no. 5, pp. 953–966, 2018. [PubMed: 30060699]

- [47]. Migueles JH, Delisle CN, Henriksson P, Cadenas-Sanchez C, Ortega FB, and Löf M, "Accelerometer data processing and energy expenditure estimation in preschoolers," *Medicine and Science in Sports and Exercise*, vol. 51, no. 3, pp. 590–598, 2019. [PubMed: 30303935]
- [48]. Chen KY and Sun M, "Improving energy expenditure estimation by using a triaxial accelerometer," *Journal of Applied Physiology*, vol. 83, no. 6, pp. 2112–2122, 1997. [PubMed: 9390989]
- [49]. Morris CE, Wessel PA, Tinius RA, Schafer MA, and Maples JM, "Validity of activity trackers in estimating energy expenditure during high-intensity functional training," *Research Quarterly for Exercise and Sport*, pp. 1–8, 2019.
- [50]. Cvetkovi B, Milic R, and Lutrek M, "Estimating energy expenditure with multiple models using different wearable sensors," *Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1081–1087, 2015.
- [51]. Altini M, Penders J, Vullers R, and Amft O, "Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning," *Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 219–226, 2014.
- [52]. Reeve MD, Pumpa KL, and Ball N, "Accuracy of the sensewear armband mini and the bodymedia fit in resistance training," *Journal of Science and Medicine in Sport*, vol. 17, no. 6, pp. 630–634, 2014. [PubMed: 24120404]
- [53]. Gjoreski H, Kaluza B, Gams M, and Milic, "Context-based ensemble method for human energy expenditure estimation," *Applied Soft Computing*, vol. 37, pp. 960–970, 2015.
- [54]. Wang J-S, Chuang F-C, and Yang Y-T, "An energy expenditure estimation algorithm for a wearable system," *International Conference of Health Informatics and Medical Systems*, p. 185, 2015.
- [55]. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, and Marshall S, "A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers," *Physiological Measurement*, vol. 35, no. 11, p. 2191, 2014. [PubMed: 25340969]
- [56]. Hedegaard M, Anvari-Moghaddam A, Jensen BK, Jensen CB, Pedersen MK, and Samani A, "Prediction of energy expenditure during activities of daily living by a wearable set of inertial sensors," *Medical Engineering and Physics*, vol. 75, pp. 13–22, 2020. [PubMed: 31679905]
- [57]. LaMunion SR, Blythe AL, Hibbing PR, Kaplan AS, Clendenin BJ, and Crouter SE, "Use of consumer monitors for estimating energy expenditure in youth," *Applied Physiology, Nutrition, and Metabolism*, vol. 45, no. 2, pp. 161–168, 2020.
- [58]. St-Laurent A, Mony M, Mathieu M, and Ruchat S, "Validation of the fitbit zip and fitbit flex with pregnant women in free-living conditions," *Journal of medical engineering & technology*, vol. 42, no. 4, pp. 259–264, 2018. [PubMed: 30198806]
- [59]. Torii M, Yamasaki M, Sasaki T, and Nakayama H, "Fall in skin temperature of exercising man," *British Journal of Sports Medicine*, vol. 26, no. 1, pp. 29–32, 1992. [PubMed: 1600450]
- [60]. Ludwig N, Formenti D, Trecroci A, Gargano M, and Alberti G, "Comparison of image analysis methods in skin temperature measurements during physical exercise," in *Quantitative InfraRed Thermography*, 2014, p. 62.
- [61]. Sevil M, Hajizadeh I, Samadi S, Feng J, Lazaro C, Frantz N, Yu X, Brandt R, Maloney Z, and Cinar A, "Social and competition stress detection with wristband physiological signals," in *14th International Conference on Wearable and Implantable Body Sensor Networks IEEE*, 2017, pp. 39–42.
- [62]. Dong B, Biswas S, Montoye A, and Pfeiffer K, "Comparing metabolic energy expenditure estimation using wearable multi-sensor network and single accelerometer," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) IEEE*, 2013, pp. 2866–2869.
- [63]. Fruin ML and Rankin JW, "Validity of a multi-sensor armband in estimating rest and exercise energy expenditure," *Medicine & Science in Sports & Exercise*, vol. 36, no. 6, pp. 1063–1069, 2004. [PubMed: 15179178]

- [64]. Liu S, Gao RX, John D, Staudenmayer JW, and Freedson PS, "Multisensor data fusion for physical activity assessment," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 687–696, 2011. [PubMed: 22156943]
- [65]. Guidetti L, Meucci M, Bolletta F, Emerenziani GP, Gallotta MC, and Baldari C, "Validity, reliability and minimum detectable change of cosmed k5 portable gas exchange system in breath-by-breath mode," *PloS one*, vol. 13, no. 12, p. 1, 2018.
- [66]. Brisswalter J and Mottet D, "Energy cost and stride duration variability at preferred transition gait speed between walking and running," *Canadian Journal of Applied Physiology*, vol. 21, no. 6, pp. 471–480, 1996. [PubMed: 8959313]
- [67]. Mokhlespour Esfahani MI and Nussbaum MA, "Classifying diverse physical activities using smart garments," *Sensors*, vol. 19, no. 14, p. 3133, 2019.
- [68]. Andriacchi T, Ogle J, and Galante J, "Walking speed as a basis for normal and abnormal gait measurements," *Journal of Biomechanics*, vol. 10, no. 4, pp. 261–268, 1977. [PubMed: 858732]
- [69]. Rose J, Gamble JG, Lee J, Lee R, and Haskell WL, "The energy expenditure index: a method to quantitate and compare walking energy expenditure for children and adolescents," *Journal of Pediatric Orthopedics*, vol. 11, no. 5, pp. 571–578, 1991. [PubMed: 1918341]
- [70]. Pandolf KB, Givoni B, and Goldman R, "Predicting energy expenditure with loads while standing or walking very slowly," *Army Research Inst of Environmental Medicine*, Tech. Rep, 1976.
- [71]. Schmitz KH, Treuth M, Hannan P, McMurray R, Ring KB, Catellier D, and Pate R, "Predicting energy expenditure from accelerometry counts in adolescent girls," *Medicine and Science in Sports and Exercise*, vol. 37, no. 1, p. 155, 2005. [PubMed: 15632682]
- [72]. Arif M and Kattan A, "Physical activities monitoring using wearable acceleration sensors attached to the body," *PloS One*, vol. 10, no. 7, p. 1, 2015.
- [73]. Anjum A and Ilyas MU, "Activity recognition using smartphone sensors," in *10th Consumer Communications and Networking Conference IEEE*, 2013, pp. 914–919.
- [74]. Wu W, Dasgupta S, Ramirez EE, Peterson C, and Norman GJ, "Classification accuracies of physical activities using smartphone motion sensors," *Journal of Medical Internet Research*, vol. 14, no. 5, p. 130, 2012.
- [75]. Lee S, Park H, Hong S, Lee K, and Kim Y, "A study on the activity classification using a triaxial accelerometer," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3 IEEE, 2003, pp. 2941–2943.
- [76]. Fortune E, Tierney M, Scanail CN, Bourke A, Kennedy N, and Nelson J, "Activity level classification algorithm using shimmer wearable sensors for individuals with rheumatoid arthritis," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE*, 2011, pp. 3059–3062.
- [77]. Sekine M, Tamura T, Fujimoto T, and Fukui Y, "Classification of walking pattern using acceleration waveform in elderly people," in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2 IEEE, 2000, pp. 1356–1359.
- [78]. Wang N, Ambikairajah E, Celler BG, and Lovell NH, "Accelerometry based classification of gait patterns using empirical mode decomposition," in *International Conference on Acoustics, Speech and Signal Processing IEEE*, 2008, pp. 617–620.
- [79]. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, and Yang Y, "Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: A systematic review," *Journal of Biomedical Informatics*, p. 1, 2018.
- [80]. Ingraham KA, Ferris DP, and Remy CD, "Using wearable physiological sensors to predict energy expenditure," in *2017 International Conference on Rehabilitation Robotics IEEE*, 2017, pp. 340–345.
- [81]. Gjoreski M, Janko V, Gjoreski H, Cvetkov B, Luštrek M, and Gams M, "Activity and stress monitoring using smartphone and wrist device," in *7th International Postgraduate School Students Conference*, 2016, pp. 154–164.

- [82]. Chowdhury AK, Tjondronegoro D, Chandran V, Zhang J, and Trost SG, "Prediction of relative physical activity intensity using multimodal sensing of physiological data," *Sensors*, vol. 19, no. 20, p. 4509, 2019.
- [83]. Chowdhury AK, Tjondronegoro D, Zhang J, Pratiwi PS, and Trost SG, "Towards non-laboratory prediction of relative physical activity intensities from multimodal wearable sensor data," in *Life Sciences Conference (LSC)*. IEEE, 2017, pp. 230–233.
- [84]. Xia V, Jaques N, Taylor S, Fedor S, and Picard R, "Active learning for electrodermal activity classification," in *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* IEEE, 2015, pp. 1–6.
- [85]. Cvetkovi B, Janko V, Gradišek A, Luštrek M, Kajtna T, and Štrumbelj B, "Mobile application to stimulate physical activity in schoolchildren," in *12th International Conference on Intelligent Environments (IE) IEEE*, 2016, pp. 206–209.
- [86]. Majumder S, Mondal T, and Deen MJ, "Wearable sensors for remote health monitoring," *Sensors*, vol. 17, no. 1, p. 130, 2017.
- [87]. Chowdhury AK, "Sensor-based prediction of physical activity and its impacts using machine learning," Ph.D. dissertation, Queensland University of Technology, 2018.
- [88]. Parent M, Tiwari A, Albuquerque I, Gagnon J-F, Lafond D, Tremblay S, and Falk TH, "A multimodal approach to improve the robustness of physiological stress prediction during physical activity," in *International Conference on Systems, Man and Cybernetics IEEE*, 2019, pp. 4131–4136.
- [89]. Wu JK, Dong L, and Xiao W, "Real-time physical activity classification and tracking using wearable sensors," in *6th International Conference on Information, Communications & Signal Processing IEEE*, 2007, pp. 1–6.
- [90]. Sevil M, Rashid M, Askari R, Mohammad, Samadi S, Hajizadeh I, and Cinar A, "Psychological stress detection using photoplethysmography," in *International Conference on Biomedical Health and Informatics IEEE*, 2019, p. 1.
- [91]. Askari MR, Rashid M, Sevil M, Hajizadeh I, Brandt R, Samadi S, and Cinar A, "Artifact removal from data generated by nonlinear systems: Heart rate estimation from blood volume pulse signal," *Industrial & Engineering Chemistry Research*, 2019.
- [92]. Zhou G, Hansen JH, and Kaiser JF, "Nonlinear feature based classification of speech under stress," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, pp. 201–216, 2001.
- [93]. Mannini A and Sabatini AM, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010. [PubMed: 22205862]
- [94]. San-Segundo R, Montero JM, Barra-Chicote R, Fernández F, and Pardo JM, "Feature extraction from smartphone inertial signals for human activity segmentation," *Signal Processing*, vol. 120, pp. 359–372, 2016.
- [95]. Bornoïu I-V and Grigore O, "A study about feature extraction for stress detection," in *8th International Symposium on Advanced Topics in Electrical Engineering IEEE*, 2013, pp. 1–4.
- [96]. Zhai J and Barreto A, "Stress detection in computer users based on digital signal processing of noninvasive physiological variables," in *International Conference of the IEEE Engineering in Medicine and Biology Society IEEE*, 2006, pp. 1355–1358.
- [97]. Bastos-Filho TF, Ferreira A, Atencio AC, Arjunan S, and Kumar D, "Evaluation of feature extraction techniques in emotional state recognition," in *4th International Conference on Intelligent Human Computer Interaction IEEE*, 2012, pp. 1–6.
- [98]. Trost SG, Wong W-K, Pfeiffer KA, and Zheng Y, "Artificial neural networks to predict activity type and energy expenditure in youth," *Medicine and Science in Sports and Exercise*, vol. 44, no. 9, p. 1801, 2012. [PubMed: 22525766]
- [99]. Sevil M, Rashid M, Hajizadeh I, Maloney Z, Samadi S, Askari MR, Brandt R, Hobbs N, Park M, Quinn L et al., "Assessing the effects of stress response on glucose variations," in *16th International Conference on Wearable and Implantable Body Sensor Networks (BSN) IEEE*, 2019, pp. 1–4.

- [100]. He H, Bai Y, Garcia EA, and Li S, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) IEEE, 2008, pp. 1322–1328.
- [101]. Jolliffe I, Principal component analysis. Springer, 2011.
- [102]. Al Shalabi L and Shaaban Z, "Normalization as a preprocessing engine for data mining and the approach of preference matrix," in International Conference on Dependability of Computer Systems IEEE, 2006, pp. 207–214.
- [103]. Jain A, Nandakumar K, and Ross A, "Score normalization in multimodal biometric systems," Pattern Recognition, vol. 38, no. 12, pp. 2270–2285, 2005.
- [104]. Hwang W-J and Wen K-W, "Fast knn classification algorithm based on partial distance search," Electronics Letters, vol. 34, pp. 2062–2063, 11 1998.
- [105]. Scholkopf B and Smola AJ, Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2001.
- [106]. Friedl MA and Brodley CE, "Decision tree classification of land cover from remotely sensed data," Remote Sensing of Environment, vol. 61, no. 3, pp. 399–409, 1997.
- [107]. Ng AY and Jordan MI, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in Advances in Neural Information Processing Systems, 2002, pp. 841–848.
- [108]. Specht DF, "A general regression neural network," IEEE Transactions on Neural Networks, vol. 2, no. 6, pp. 568–576, 1991. [PubMed: 18282872]
- [109]. Heaton J, Introduction to neural networks with Java. Heaton Research, 2008.
- [110]. Rasmussen CE, "Gaussian processes in machine learning," in Machine Learning. Springer, 2003, pp. 63–71.
- [111]. Dietterich TG et al., "Ensemble learning," The Handbook of Brain Theory and Neural Networks, vol. 2, pp. 110–125, 2002.
- [112]. Graybill FA, Theory and application of the linear model. Duxbury Press North Scituate, 1976.
- [113]. Lecun Y, Bengio Y, and Hinton G, "Deep learning," Nature, vol. 521, no. 7553, p. 436, 2015. [PubMed: 26017442]
- [114]. Grushin A, Monner DD, Reggia JA, and Mishra A, "Robust human action recognition via long short-term memory," in International Joint Conference on Neural Networks IEEE, 2013, pp. 1–8.
- [115]. Greff K, Srivastava RK, Koutník J, Steunebrink BR, and Schmidhuber J, "Lstm: A search space odyssey," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222–2232, 2016. [PubMed: 27411231]
- [116]. Snoek J, Larochelle H, and Adams RP, "Practical bayesian optimization of machine learning algorithms," in Advances in Neural Information Processing Systems, 2012, pp. 2951–2959.
- [117]. Montoye AH, Pivarnik JM, Mudd LM, Biswas S, and Pfeiffer KA, "Comparison of activity type classification accuracy from accelerometers worn on the hip, wrists, and thigh in young, apparently healthy adults," Measurement in Physical Education and Exercise Science, vol. 20, no. 3, pp. 173–183, 2016.
- [118]. Wundersitz DW, Josman C, Gupta R, Netto KJ, Gastin PB, and Robertson S, "Classification of team sport activities using a single wearable tracking device," Journal of Biomechanics, vol. 48, no. 15, pp. 3975–3981, 2015. [PubMed: 26472301]
- [119]. el Achkar CM, Lenoble-Hoskovec C, Paraschiv-Ionescu A, Major K, Büla C, and Aminian K, "Instrumented shoes for activity classification in the elderly," Gait and Posture, vol. 44, pp. 12–17, 2016. [PubMed: 27004626]
- [120]. Moncada-Torres A, Leuenberger K, Gonzenbach R, Luft A, and Gassert R, "Activity classification based on inertial and barometric pressure sensors at different anatomical locations," Physiological Measurement, vol. 35, no. 7, p. 1245, 2014. [PubMed: 24853451]
- [121]. Das B, Cook DJ, Krishnan NC, and Schmitter-Edgecombe M, "One-class classification-based real-time activity error detection in smart homes," Journal of Selected Topics in Signal Processing, vol. 10, no. 5, pp. 914–923, 2016.

- [122]. Mansoubi M, Pearson N, Clemes SA, Biddle SJ, Bodicoat DH, and Tolfrey, "Energy expenditure during common sitting and standing tasks: examining the 1.5 met definition of sedentary behaviour," BMC Public Health, vol. 15, no. 1, p. 516, 2015. [PubMed: 26021449]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.
Raw and Processed BVP Signals

Fig. 2.

Example of the Time Delay in GSR Measurements Relative to the Start of Exercise as ACC and EE Rise in Advance of the GSR Readings

Fig. 3.
Flow Chart of Proposed Deep Learning Structure

Fig. 4.
Structure of LSTM Model

Fig. 5.

Accuracy of Different Machine Learning Algorithms (Testing Data Set (10%)) (EL: Ensemble Learning, LSTM-DL: Long-Short Term Memory-Deep Learning, k-NN: k Nearest Neighbours, LD: Linear Discrimination, SVM: Support Vector Machine, DT: Decision Tree)

Fig. 6.
ROC with LSTM Algorithm

Output Class	SS	1668 16.2%	0 0.0%	4 0.0%	0 0.0%	21 0.2%	98.5% 1.5%
	DA	238 2.3%	1842 17.9%	0 0.0%	0 0.0%	0 0.0%	88.6% 11.4%
	RT	0 0.0%	89 0.9%	2261 22.0%	0 0.0%	0 0.0%	96.2% 3.8%
	RN	0 0.0%	0 0.0%	123 1.2%	1763 17.1%	0 0.0%	93.5% 6.5%
	BK	0 0.0%	0 0.0%	32 0.3%	27 0.3%	2225 21.6%	97.4% 2.6%
		87.5% 12.5%	95.4% 4.6%	93.4% 6.6%	98.5% 1.5%	99.1% 0.9%	94.8% 5.2%
		SS	DA	RT	RN	BK	
Target Class							

Fig. 7.

Confusion Matrix for LSTM-DL based Classification Result

Fig. 8.
EE Estimation Results Using an Independent Testing Dataset

TABLE I

Biosensors of Empatica E4, Biosignals and Example Indicators

Sensor	Biosignal	Example Indicator
Infrared Thermopile	ST	Increased with resistance exercise [60], fall during biking exercise [59]
PPG	BVP	Alteration during bike exercise, noise (motion artifact) increase during any kind of physical activity
3-D ACC	ACC	Intensity of running, relatively stable during sedentary state and bike
PPG	HR	Increased during running, bike and resistance exercise, stable during sedentary state
GSR	GSR	Capturing sweating rate, intensity of daily activities, bike or running

TABLE II

Breakdown of the PS and EE Experiments for Data Collection

PS Experiments with Empatica E4			
PA	Number Experiments	Number Subjects	Total Time (min)
SS	96	7	16758
DA	43	7	2783
RT	24	8	1016
RN	108	21	3741
BK	56	16	1602
EE Experiments with Empatica E4 and COSMED K5			
PA	Number Experiments	Number Subjects	Total Time (min)
SS	10	4	2432
DA	6	4	673
RT	9	2	296
RN	30	7	645
BK	43	10	791

TABLE III

List of Select Informative Features

Physical State Classification		
Feature Name	Time Delay (min)	p-value
Norm of 4th detail coefficient of Daubechies wavelet decomposition of ACC X-axis	0	0.01
Total energy response of BVP	1	0.01
Range for Magnitudes of ACC	0	0.01
Kurtosis of GSR	3	0.01
Standard Deviation of ACC / Std of GSR	0	0.01
Median of HR / Median of GSR	0	0.01
Std of ACC / Mean of ACC	0	0.01
Skewness of BVP	1	0.01
Minimum of HR	0	0.01
Mean of ST	0	0.01
p-values calculated using a two-sided t-test at $\alpha = 1\%$ significance level		
EE Estimation (Features with COSMED EE Measurements)		
Feature Name	Time Delay (min)	Correlation
Std of magnitude of ACC (X-Y-Z)	0	0.56
Mean of logarithm of magnitude of ACC (X-Y-Z)	0	0.55
Maximum of ACC Z-axis	0	0.55
Zero cross of BVP	0	0.49
Minimum of high frequency of ST (0.15–0.4 Hz)	0	0.49
Root-mean-square energy of GSR	4	0.48
Maximum of power spectral density estimate of HR	0	0.48
1st Quartile of HR	0	0.48
Norm of 1st detail coefficient of Daubechies wavelet decomposition of GSR	0	0.48
Summation of absolute differentiation of GSR	12	0.34

TABLE IV

List of Hyperparameters Optimized for the ML Algorithms

ML Algorithm	Hyperparameter(s)
SVM/SVR	Kernel Scale, C parameter
DT	Min. Leaf Size
k-NN	Num. of Neighbours, Distance Function
LD	Delta, Gamma
NB	Distribution Name, Width
EL	Method, Learning Rate, Minimum Leaf Size

TABLE V

Recall, Precision and F-Score (SS: Sedentary State, DA: Daily Activity, RT: Resistance Training, RN: Running, BK: Biking)

Recall						
Algorithm	SS	DA	RT	RN	BK	Mean
LSTM DL	0.875	0.954	0.934	0.985	0.991	0.947
k-NN	0.918	0.865	0.896	0.976	0.985	0.928
SVM	0.820	0.805	0.843	0.908	0.890	0.853
LR	0.824	0.666	0.822	0.922	0.851	0.817
NB	0.930	0.592	0.840	0.711	0.822	0.779
DT	0.898	0.837	0.908	0.896	0.889	0.886
EL	0.973	0.901	0.934	0.960	0.907	0.935

Precision						
Algorithm	SS	DA	RT	RN	BK	Mean
LSTM DL	0.985	0.886	0.962	0.935	0.974	0.948
k-NN	0.968	0.936	0.985	0.837	0.890	0.923
SVM	0.906	0.767	0.884	0.864	0.852	0.855
LR	0.852	0.796	0.833	0.758	0.797	0.807
NB	0.836	0.792	0.703	0.755	0.727	0.763
DT	0.903	0.848	0.914	0.884	0.879	0.886
EL	0.956	0.924	0.957	0.876	0.943	0.931

F-score						
Algorithm	SS	DA	RT	RN	BK	Mean
LSTM DL	0.927	0.919	0.948	0.959	0.982	0.947
k-NN	0.942	0.899	0.938	0.901	0.935	0.923
SVM	0.861	0.785	0.863	0.886	0.871	0.853
LR	0.837	0.726	0.827	0.832	0.823	0.809
NB	0.881	0.678	0.766	0.732	0.771	0.766
DT	0.901	0.843	0.911	0.890	0.884	0.886
EL	0.964	0.912	0.946	0.916	0.925	0.933

TABLE VI

Comparison of PS Classification and EE Estimation Results of Our Proposed Method with Other Similar Studies (RF: Random Forest; LMT: Logistic Model Tree; Inertial sensor includes 3D sensing unit: accelerometers, magnetometers, and gyroscopes; All accelerometer sensors were 3-D)

Comparison of PS Classification Results						
Year	Reference	Used Sensors	Accuracy(%)	Algorithm	Sensor Place	Device
2016	[117]	ACC	81	NN	Wrist	Wrist-mounted 3-D Accelerometers
2015	[118]	ACC, Gyroscope	85	LMT	Neck	Minimax S4
2016	[119]	Inertial Sensor	97	DT	Shoes	Physilog Gait Up
2015	[120]	Inertial Sensor	95	k-NN	Wrist, Ankle, Trunk	ReSense
2020	[35]	ACC	95	RF	Hip	ActiGraph
2020	[38]	ACC	93	DL	—	Smart Phone
2020	[37]	ACC	89	SVM	Hip and Thigh	GT3X ₊ and GT9X
2020	[36]	ACC, BVP, HR	92	RF	Hip and Wrist	Microsoft Band, ADXL335
—	Proposed Method	ACC, HR, ST, GSR, BVP	95	LSTM-DL	Wrist	Empatica E4
—	Proposed Method	ACC (Using only ACC)	90	LSTM-DL	Wrist	Empatica E4

Comparison of EE Estimation Results						
Year	Reference	Used Sensors	MAE (MET)	MAPE (%)	RMSE (MET)	Device
2015	[50]	ACC, HR, NB	0.52	23.2	—	Zephyr, Shimmer2, BodyMedia Fit
2014	[51]	ACC, ECG	0.70	25.7	—	ADXL330 and five ECG Necklace
2015	[52]	ACC, HR, ST	0.87	32.8	—	BodyMedia FIT
2015	[53]	ACC, HR, RR	0.61	27.5	—	BodyMedia Fit, Zephyr, Shimmer
2015	[54]	ACC	0.31	—	—	LIS3L02AQ3 Sensor Data
2014	[55]	ACC, HR	—	—	1.0	ActiGraph GT3X ₊ , Polar RS400
2020	[56]	HR	—	—	0.71	Suunto Smart Belt
2020	[57]	HR	—	16.8	—	Misfit Shine 2
—	Proposed Method	ACC, HR, ST, GSR, BVP	0.25	12.1	0.51	Empatica E4
—	Proposed Method	ACC (Using only ACC)	0.59	22.0	0.79	Empatica E4

TABLE VII

Comparison of EE Estimation with Different Machine Learning Algorithms and Metrics

Algorithm	RMSE (MET)	MAPE (%)	MAE (MET)
LSTM DL	0.509	12.086	0.254
k-NN	0.520	14.763	0.335
SVM	0.537	13.174	0.264
LR	0.533	12.954	0.262
NB	0.610	15.603	0.464
GPR	0.534	12.880	0.261
EL	0.608	14.029	0.437
DT	1.539	25.157	0.961