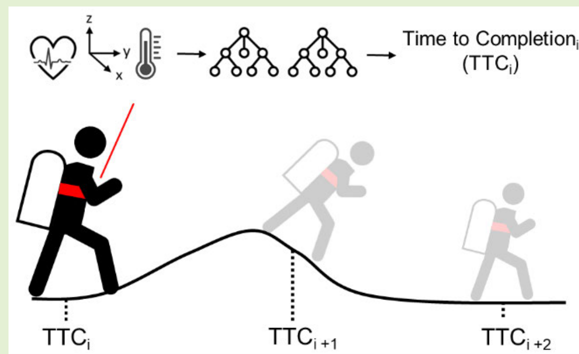
Predicting Soldier Performance on Structured Military Training Marches With Wearable Accelerometer and Physiological Data

David J. Lin, *Graduate Student Member, IEEE*, Aprameya Satish, *Member, IEEE*, Kristine L. S. Richardson, *Graduate Student Member, IEEE*, Sungtae An, Cem O. Yaldiz, Mark Buller, Kyla Driver, Emma Atkinson, Timothy Mesite, Christopher King, Omer T. Inan, *Senior Member, IEEE*, and Alessio Medda, *Senior Member, IEEE*

***Abstract*—Performance metrics are extremely important for military instructors and leadership to gauge soldier fitness and readiness and adjust training regimens accordingly. One important performance indicator during military training events is the time taken for a soldier to complete a structured march, hereon referred to as time-to-completion (TTC). During these marches, wearable physiological sensors can be used to infer a soldier's physiological state and exertion rate, which in turn can be used for predicting TTC. In this work, we present a model that uses signals from a multimodal wearable sensor to predict TTC for soldiers undergoing a 12-mile structured ruck march. Predictions are made at discrete time points (checkpoints) throughout the march using features from skin temperature (SKT), heart rate (HR), estimated core temperature (ECT), and triaxial accelerometry. To utilize the structured nature of these marches, separate models are trained at each checkpoint using features from both the current and past checkpoints. By 120 min (2/3 of the expected 180-min completion time), we achieved a TTC root-mean-square error (RMSE) of 7.12 min and a mean absolute error (MAE) of 5.21 min using this model. Integral to TTC estimation accuracy were gait-related features such as the standard deviation of vertical acceleration (ACC). Features such as HR slope and performance metrics from prior exercises minimally improved accuracy. The deployment of this model will enable continuous monitoring of performance metrics for online TTC estimation.**



***Index Terms*— Military training, physical fitness, random forest regression, wearable devices.**

I. Introduction

ASSESSING soldier fitness is important to identify unit readiness during training and prior to deployment. Furthermore, individual soldier fitness assessments can inform military leadership of training regimen optimizations over personal-to-regiment levels. The United States (U.S.) Army conducts regular fitness tests [1] for soldiers that comprise a series of exercises, such as general physical training, timed runs, and timed ruck marches. These are currently monitored by manually keeping track of soldier performance through metrics such as the number of repetitions completed or time taken to complete an exercise.

Ruck marches, specifically, have been shown to be effective indicators of soldier fitness as they are conducted over challenging terrain with the soldiers carrying heavy loads in excess of 25 kg on their backs, reproducing realistic operational conditions [2]. These marches are usually timed, and soldiers are required to complete a fixed distance over a specified route within the allowed time. This article refers to such marches as "structured marches." A heuristic metric of march performance is the time taken by soldiers to complete the event or time-to-completion (TTC). At present, the army measures completion time as the soldier reaches the end of the prescribed course. With access to data from portable wearables tracking accelerometry and physiology [3],

Manuscript received 30 July 2023; accepted 20 October 2023. Date of publication 10 November 2023; date of current version 2 January 2024. This work was supported by the Office of Naval Research under Grant N00014-20-1-2137. The associate editor coordinating the review of this article and approving it for publication was Dr. Qammer H. Abbasi. *(Corresponding author: David J. Lin.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Research and Development Command Institutional Review Board under Application No. M-10720.

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/JSEN.2023.3330409

near real-time updates on soldier progress through the march become possible, and offer a more comprehensive photograph of soldier performance. Specifically, the expected TTC for a soldier at any given march instance could be provided to the individual and/or the unit commander, and based on that value, the pace during the structured march may be increased or decreased accordingly, in real-time, to maximize the probability of finishing within the goal for TTC for that individual.

The use of a global positioning system (GPS) may make estimating the TTC more accurate as it can provide better estimates of velocity than direct integration of accelerometry data can. GPS sensors can be installed in wearable devices and fit onto each soldier, enabling constant updates of TTC. However, because soldiers need to complete multiple exercises during the whole day, an unobtrusive, rugged, portable sensor with a long battery life is preferable [4], [5]; these design constraints can limit the integration of GPS into the sensing hardware from a packaging and power consumption standpoint [6]. GPS accuracy can also degrade for various environmental reasons including low signal-to-noise ratio (SNR), multipath errors, and limited line of sight to satellites when marching through regions of dense tree cover [7]. The accuracy can also be affected by operational errors including improperly wearing the sensor, device misuse, or device initialization errors [8]. These limitations of GPS make it difficult to solely rely on satellite positioning for TTC prediction. As such, the ability to estimate TTC from current sensor systems that measure accelerometry and other physiological signals locally is still desired.

To the best of the authors' knowledge, this is the first body of work seeking to estimate the TTC of a soldier for a structured march. Related works on estimating a completion time include stride length estimation using accelerometer data by Xing et al. [9], where a neural network trained on accelerometer features and participant height was used to estimate the distance covered per step. However, this study focused on a pedestrian population with no load carriage and required wearing the monitoring sensor on the foot. Furthermore, in this study, soldier heights were not readily available, making direct stride length estimation using Xing et al.'s [9] method infeasible. Moreover, information from physiological measurements was not used, and stride estimation was achieved using gyroscopes. While gyroscopes are generally useful for activity monitoring and stride length estimation [10], [11], this sensing modality was not available during this study.

There are, however, alternative measures that may be used to predict performance for a soldier over a structured march using only accelerometer and physiological data. For example, it is well established that heart rate (HR) can be a proportional measure of metabolic energy expenditure [9], [12] at a given point during the march, while skin and core temperature (CT) measurements can indicate physical exertion [13]. Moreover, the maximization of lateral acceleration (ACC) and the minimization of vertical ACC have been shown to improve gait efficiency in [14], suggesting the use of ACC power as an efficient indicator of performance. Studies have also shown that high-performing athletes are able to maintain lower HRs for long durations of time during endurance training [15], [16], suggesting that HR dynamics could be used as a metric of march performance. A predictive model to estimate TTC can be developed using a collection of such derived features, where the model can infer relative feature importance based on the sensor data available when making its prediction. This semiempirical approach directly handles the complex interrelationship between gait parameters and physiology per individual over the march.

In this work, we present a model, the first to our knowledge that uses physiological measurement [skin temperature (SKT), HR, and estimated CT (ECT)] and triaxial accelerometry from wearable sensors to accurately predict TTC over a structured march. The total length of the course is divided into checkpoints, and at each checkpoint, a new prediction of the TTC is made. A random forest regressor at each checkpoint is trained to provide this new TTC prediction using features not only from the current checkpoint but also from previous checkpoints in the march. The model is trained and tested on a large population of soldiers performing a structured march.

To summarize, our work contributes to the existing literature by introducing a physiology-driven TTC prediction model that serves as a metric for structured march performance.

II. Methods

A. Experimental Protocol

This study used data collected under a protocol approved by the Medical Research and Development Command Institutional Review Board (Protocol number M-10720). These data were collected during 12-mile loaded rucksack marches and 5-mile runs during the Ranger Assessment and Selection Program (RASP) at Fort Moore, Columbus, GA, USA. Soldiers were included in the analysis if they participated in both exercises and if their 12-mile march TTC was within an acceptable range. The latter is explained further in detail in Section II-D. The study was comprised predominantly of male soldiers (23 $\pm$ 4) years of age, with an average height of (1.77 $\pm$ 0.08) m, and body weight (78.1 $\pm$ 10.4) kg. Each subject participated in a single march event.

Data from 468 soldiers who underwent both a 12-mile march on a predetermined route and a 5-mile run on a paved track were used. Soldiers were expected to complete the 12-mile march within a 180-min timeframe and were free to use their favorite combination of military marching and running strategies. Soldiers were equipped with their army combat uniform and carried a loaded rucksack and a weapon totaling 14 kg in combined weight. For the 5-mile run, soldiers were expected to complete the exercise in 40 min. Soldiers did not carry a load during this exercise.

During each exercise, participants were equipped with a custom torso-worn wearable sensor [Open Body Area Network Physiological Status Monitor (OBAN-PSM)] strapped around the chest to measure and log HR, SKT, and ACC signals [17]. HR was captured using a custom dry electrode electrocardiogram (ECG) sensor around the chest. Although the ECG is sampled at a much higher rate (512 Hz), the sensor does not store the raw ECG waveform and instead outputs a time-averaged HR estimate every 5 S (0.2 Hz). Triaxial

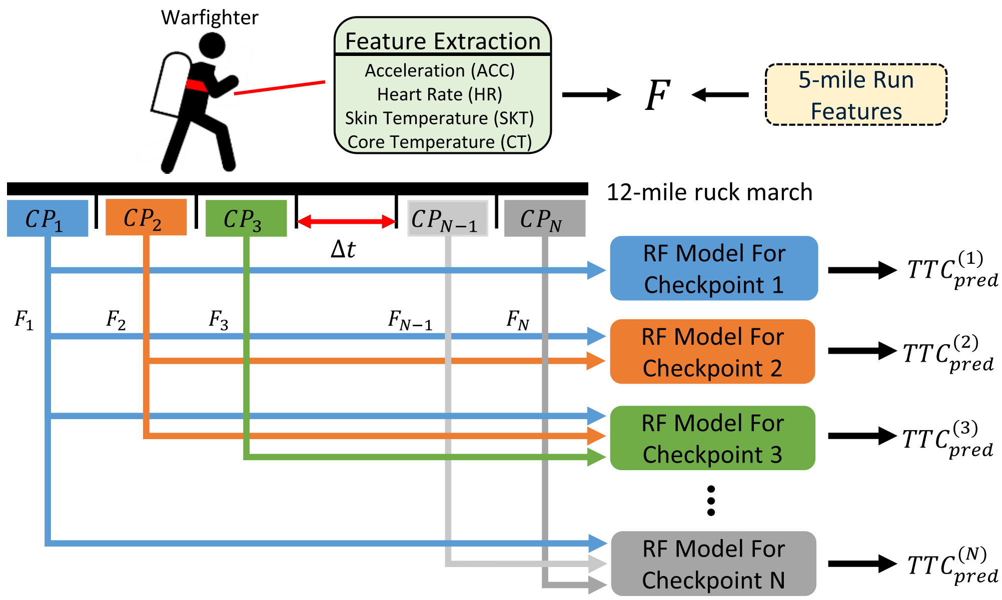


Fig. 1. Model architecture for TTC estimation. The overall duration 12-mile ruck march for each subject is divided into equal segments of length Δt called checkpoints (***CP***). Signals within each checkpoint (CP_i) are then used to calculate features and subsequently combined with 5-mile run derived features to form the feature vector F_i for CP_i. The associated model at CP_i is then trained on the feature set [F_i, F_{i-1}, F_1] corresponding to features from the current, previous, and baseline checkpoints to estimate the TTC at that point in time. In this work, Δt is set to 10 min.

ACC was captured using an accelerometer (ADXL362 chip, ±8 g; Analogue Devices, Norwood, MA, USA) and sampled at 128 Hz. SKT was sampled at 0.5 Hz. CT was subsequently derived using a Kalman-filter architecture applied to the heart rate measurements known as the ECTemp algorithm developed by Buller et al. [13].

B. Model Architecture Overview

Fig. 1 shows the model architecture overview used to estimate TTC. The 12-mile ruck march duration was first divided into standard epochs. In this work, we refer to these epochs as "checkpoints" (CP), which serve as reference gates for predicting TTC. For each subject, features F_i for each checkpoint CP_i were formed using both the signals measured at that checkpoint and from the associated 5-mile run. These "global" 5-mile run features are assumed to provide previous knowledge of soldier fitness. As previously mentioned, subjects who did not have a 5-mile run were not included in the current analysis. In this work, a 10-min-long epoch was chosen as the checkpoint length to provide TTC estimates at approximately every 5% of the total expected march duration. A parametric study of checkpoint length was conducted, and a 10-min checkpoint length was found to offer a reasonable compromise between computational cost and prediction accuracy, with shorter checkpoints performing similarly, and longer checkpoints reducing prediction accuracy.

TTC was then estimated at each checkpoint by inputting features from multiple checkpoints into a distinct random forest regression model [18]. Specifically, the features from the current checkpoint (F_i), previous checkpoint (F_{i-1}), and baseline checkpoint (F_1) were concatenated together and used as an input to the corresponding model. Here, the baseline feature matrix is the feature matrix before the first checkpoint.

Random forest regression was chosen for its ability to explain nonlinear relationships in an explainable model using feature importance [19]. For this work, we used the Gini importance index to estimate the relative importance of each feature to TTC estimation performance [18].

C. Feature Extraction

Fig. 2 shows a more detailed explanation on how features are extracted from the 12-mile ruck march and 5-mile run data. For the 12-mile march, each 10-min checkpoint was partitioned into nonoverlapping windows of 30, 10, and 15 s for HR, ACC, and SKT, respectively. HR window lengths were longer in order to have sufficient samples due to the lower sampling rate. For ACC specifically, 10-s windows were used to provide computationally efficient, but localized estimates of gait parameters, and meet the minimum prescribed window size of approximately 5 s from [20] for accurately estimating step count. A 15-s window for SKT measurements was selected to account for the lower sampling rate of SKT. A total of 21 HR features, 58 ACC features, and one temperature feature were tracked over all available windows. A single checkpoint feature was calculated as the average of features across all windows contained within the checkpoint.

The features from the ACC signals included statistical measures (mean, standard deviation, and kurtosis) and frequency band powers for each of the three axes. Approximate entropy and multiscale wavelet coefficients [21] as well as frequency power features were also collected. Notably, the vertical ACC axis was used to calculate gait-related parameters, such as cadence, step count, step time, coefficient of variation, step regularity, stride regularity, step symmetry, and step asymmetry [22]. Features for the heart rate signal included heart rate mean and frequency power band features. In addition,

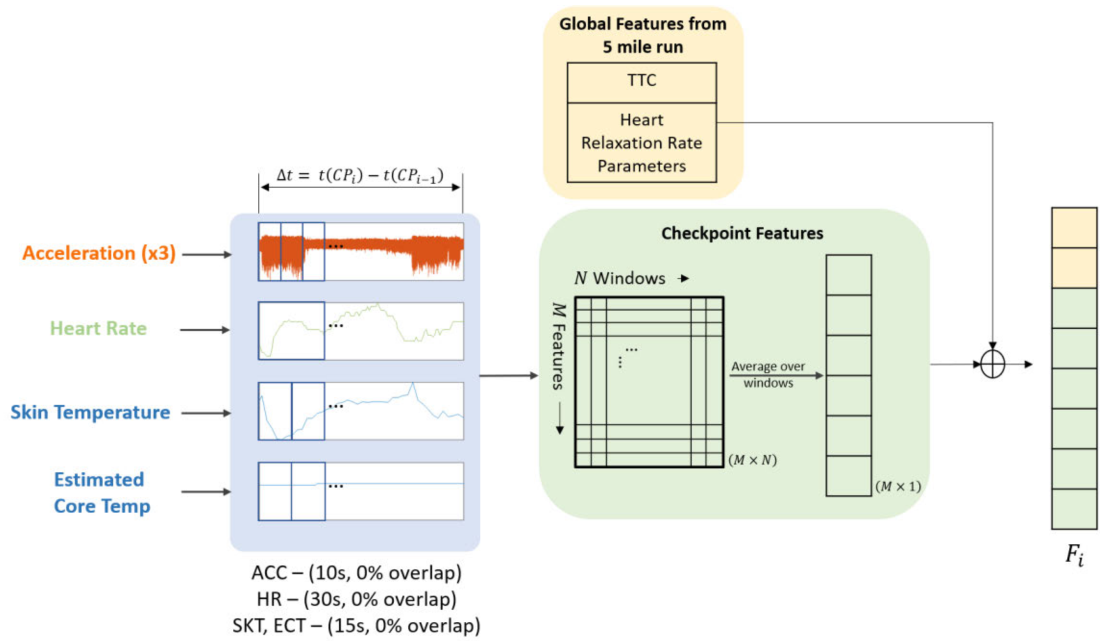


Fig. 2. Feature extraction pipeline. For each subject, signals are segmented into nonoverlapping windows (shown as blue boxes in sensor traces). Derived features (see Section II-B) are extracted for each window. For checkpoint (***CP**$_i$*), features computed for windows between the interval Δt are averaged to form the checkpoint feature matrix. The final feature matrix F_i for a subject was formed by appending the checkpoint features to the global features extracted from a 5-mile run for that subject.

HRV metrics such as the Poincare parameters were also computed [14]. Heart rate slope was also calculated as the change in average heart rate between successive windows over the window length to provide a measure of heart rate dynamics. Finally, the core and SKT difference were calculated as a metric of heat strain compensation [23].

To incorporate prior knowledge of soldier fitness, features from each soldier's corresponding 5-mile run were also taken into consideration. Specifically, the TTC for the 5-mile run (TTC—5 m) and the time constants from the postexercise heart rate recovery models [24], [25] were added as global features. This global TTC of 5 m was chosen as a baseline of a subject's general performance, while the heart rate recovery constants were included after the work of Pierpont et al. [24], which demonstrated that they could be used as an index of sympathetic withdrawal and parasympathetic reactivation after strenuous activity. These features were appended onto the 12-mile march feature matrix for each subject.

D. TTC Label Annotation

Since GPS information was not available for the data used in this study, TTC was manually calculated for each soldier using the raw vertical ACC and HR signals for both the 5-mile run and 12-mile march. Time boundaries were visually extracted using changes in energy between minimal activity (e.g., standing) and exertion at the start and end of both the march and run. The TTC for a given activity was then computed as the difference between these boundaries. The final TTC for a participant was then calculated by taking the average of the ACC and heart rate-derived TTCs. Soldiers with a TTC of less than 120 min were removed from the analysis. This threshold was set because a TTC of less than 120 min can be considered to be physically unrealistic for the 12-mile march. Indeed, the subjects that were removed based on this threshold were either those whose sensor data were corrupted or who were extracted from the field for medical reasons (one subject suffered from heat stroke). This resulted in the final subject count of ($N = 468$).

E. Random Forest Regression Training

Random forest training and testing were done using a 75/25 train-test split, randomized by subject. Training and testing splits were consistent across each checkpoint model to ensure the same subjects were always in the same testing or training sets. In addition, each model was only trained on subjects who had not finished at that checkpoint. The model at each checkpoint had an ensemble of 200 trees, trained with bagging. An inner loop threefold cross validation was used to fit hyperparameters such as maximum depth ranging from 10 to 100 and a maximum number of features to split on [18].

F. Justification for Model Design Choices

In this work, we also investigated the effect of two parameters integral to the model architecture: 1) the addition of historic feature information and 2) the use of a distinct model for each checkpoint. Their implications in the context of TTC estimation performance are explored later in Section III-B.

The additional feature matrices from the previous and baseline checkpoints were included to provide history on the extracted features. These additional features are expected to provide context to the current features such as a period of recovery after high strain. These learned contexts are expected to improve TTC estimation.

A unique random forest model was used for each checkpoint to leverage the structured nature of the exercise and provide context to acquired features. This is particularly important to identify whether regions of high strain are due to the terrain being physically draining or due to increased performance of subjects with low TTC. In addition, important features that predict TTC might evolve throughout the exercise, necessitating the need to re-evaluate the model.

III. Results and Discussion

A. Prediction Error

For all tests conducted, we analyzed performance using four different TTC prediction methods: 1) estimates using the mean of the TTC labels; 2) estimates using a model based on cadence and stride length model; 3) estimates using the proposed RF regression model with only ACC features provided; and 4) estimates using the proposed RF regression model with all features provided. We refer to the method using the mean TTC for TTC estimation as the mean TTC method and the method using cadence as the Cadence method.

The cadence-based model estimates TTC at each checkpoint using (1) with cadence and stride length as a velocity surrogate (V_i). The distance already traveled by a soldier was calculated by multiplying the velocities at all prior checkpoints and the duration of each checkpoint (Δt). These two components were used to estimate the remaining time to complete the march. The elapsed time ($i\Delta t$) was then added to output the final TTC. The total march distance (D_{tot}) was 12 miles in this work

$$\text{TTC}_i = \frac{D_{\text{tot}} - \Delta t \Sigma_k^i (V_k)}{V_i} + i\Delta t. \tag{1}$$

Cadence was calculated at each checkpoint; however, because stride length estimation has not been well-researched without gyroscope measurements [9], [11], a parameter study was conducted to find the optimal stride. Specifically, stride lengths from 75 to 90 cm were used to estimate TTC on the training set. In this work, a stride length of 86 cm resulted in the lowest root-mean-square error (RMSE), which was very similar to the average stride length of an average military soldier [26]. This was then used as the stride length on the test set.

The comparison of the first two models was conducted to investigate whether any significant prediction changes occur from the addition of physiological features. The cadence-based model was used to compare the proposed work with a simple velocity-based TTC estimation model. We only show TTC estimations up to the 120-min mark (i.e., 66% of the permitted march time) to highlight model performance early into the march. This time limit matched the lowest TTC in the data (fastest subject) and ensured that all TTC models contained the same number of test subjects. Specifically, at checkpoints past this time limit, the TTC labels being estimated naturally have a larger mean and lower standard deviation as soldiers finish the march. As such, comparisons were limited to before this time limit to ensure all models were evaluated using the same TTC label distribution.

TABLE I
Median (Mean) Error of |True TTC—Estimated TTC| Using Various Models

Time of Prediction [mins]	TTC using Mean TTC [mins]	TTC using Cadence based estimate [mins]	TTC using ACC Features [mins]	TTC using All Features [mins]
10	7.45 (8.95)	26.57 (25.21)	6.25 (7.35)	5.50 (7.19)
40	7.45 (8.95)	7.57 (8.76)	4.98 (6.18)	5.13 (6.15)
80	7.45 (8.95)	7.08 (8.33)	3.85 (5.86)	4.03 (5.78)
120	7.45 (8.95)	7.01 (8.13)	4.41 (5.74)	4.03 (5.28)

Fig. 3(a) shows the RMSE between the predicted TTC estimates and ground truth TTC labels for all tested models. While the RMSE error using the mean TTC-based method remained at ~11.70 min, the RMSE when using the RF model is lower, steadily decreasing, at 120 min, to 7.07 min when using all features and 7.73 min when using just ACC features. The addition of the physiological features slightly reduces mean TTC RMSE with a maximum difference of 0.8 min, suggesting some benefit of including these modalities for this task. The cadence-based model also shows better performance than when using the mean TTC at later checkpoints. However, the proposed model still performed better, suggesting that there is additional information in the ACC and physiological signals for better estimating TTC.

The cadence-based model had an RMSE of 27.83 min at CP_1[obscured in Fig. 3(a)], a value much greater than the RMSE at other checkpoints. The high RMSE was determined to be due to a larger estimated cadence compared with the other checkpoints. This coupled with the assumption that the estimated velocity was assumed to remain constant for the remainder of the march caused TTC_1 calculations to heavily underestimate the true TTC.

Fig. 3(b) shows the box plots for the absolute TTC error with absolute error means and medians listed in Table I for a few prediction times. As the exercise progresses, the mean and median absolute error of both RF regressors decreases. From the box plots, we can see that the upper whisker and the upper quartile for the RF with all features included is generally lower than its RF with only ACC features counterpart. However, this relationship was determined to not be significant.

Fig. 4 shows the modified Bland-Altman and correlation plots for estimated TTC using both RF configurations. TTC estimates at 120 min were included. We can see that errors are relatively symmetrical with a slight negative skewness at lower TTC values. Subjects with the fastest (<130 min) and slowest (>180 min) completion times had the highest prediction error, potentially attributed to the relatively small number of training examples that completed the march in these times. Although the 95% limits of agreement may seem high, it corresponds to an error of <10% for a total expected march time of 180 min. Furthermore, the high correlation between the estimated and true TTC suggests that although TTC cannot be precisely estimated, our model can still stratify low and high performing subjects, which is also important for leadership feedback on training effectiveness.

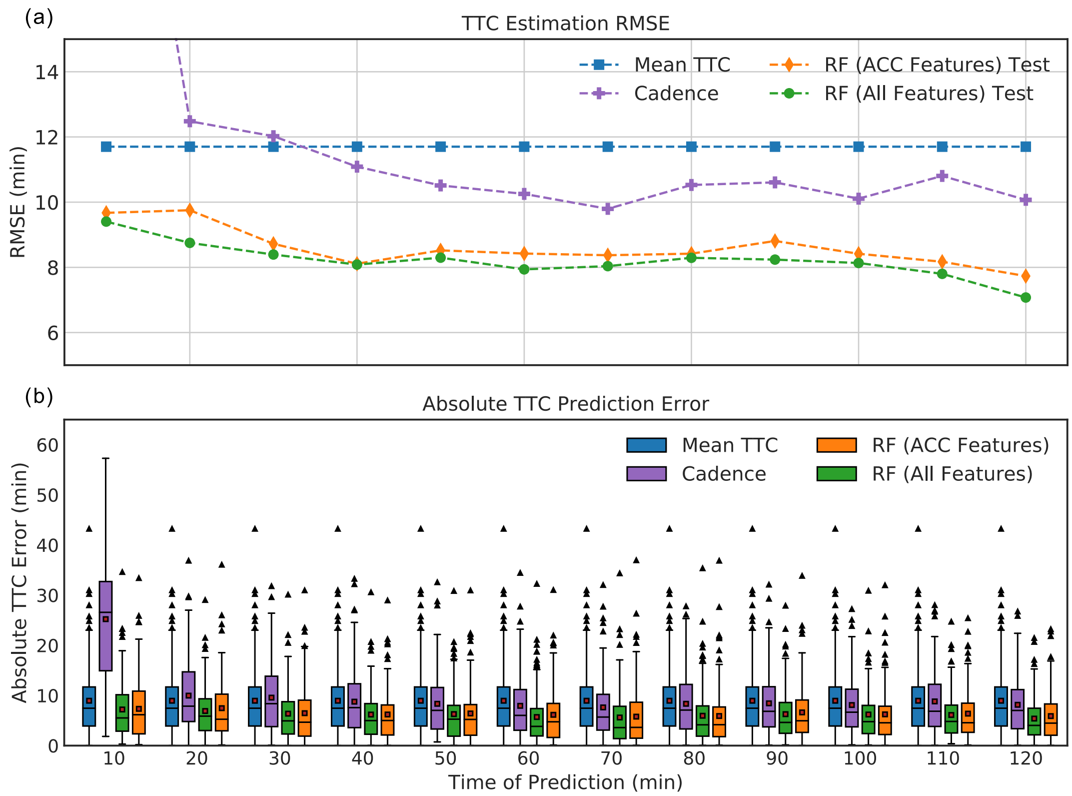


Fig. 3. (a) TTC estimate RMSE over time. Several configurations were tested: 1) estimating the mean global TTC—mean TTC; 2) estimating using a Cadence and stride length based model—Cadence; 3) estimating using the proposed model with only ACC features—RF (ACC Features); and 4) estimating using the proposed model with all features—RF (all features). For better visualization, RMSE plot ranges exclude the 10-min RMSE value when using the cadence model (27.83 min). (b) Box plot of absolute TTC estimation error when using the previously mentioned models.

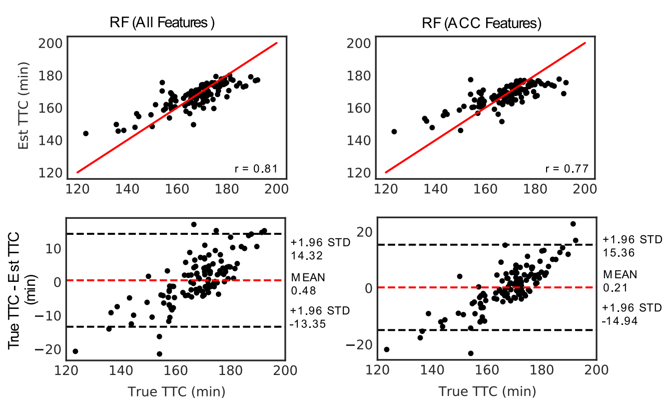


Fig. 4. Correlation (top row) and modified Bland-Altman plots between RF regression TTC estimates and true TTC estimates using ACC features or all features at 120 min. The red solid lines in the correlation plots correspond to a correlation of $r = 1$, and is shown for reference.

B. Effect of Model Data Inputs

To motivate the increased complexity of: 1) including features from previous checkpoints and 2) training an individual RF regression model for each checkpoint, additional model configurations were trained to investigate the effects of each change individually. These models were then compared with the proposed configuration of using separate RF models for each checkpoint and a feature construction using [CP_1, CP_{n-1}, CP_n] referring to features derived from the baseline, previous, and current checkpoint respectively. Models were evaluated using TTC RMSE at each checkpoint.

To test the advantages of including features from previous checkpoints for TTC estimation, RF models were trained at each checkpoint using different feature construction setups. Specifically, models were trained at each checkpoint with features from just the current checkpoint [CP_n], features from the current and previous checkpoint [CP_{n-1}, CP_n], and features from the current and baseline checkpoints [CP_1, CP_n]. We refer to these tests as multi-RF–multi-CP X, where X is the feature construction configuration.

To test the benefit of training an individual model for each checkpoint, a single model (single RF–multi-CP) was trained to predict over the entire march period. This model was trained using a feature construction of [CP_1, CP_{n-1}, CP_n] from every checkpoint. It should be noted that predictions were not made at 10 or 20 min for this model because not enough checkpoints were available to feed into the model.

Fig. 5 shows the TTC RMSE when using the additional models trained. Using the multi RF–single CP [CP_n] model configuration drastically reduced performance. However, with the inclusion of the features from the baseline checkpoint, multi-RF–multi-CP [CP_1, CP_n] showed improved performance at each checkpoint. TTC RMSE is further improved when including just the previous checkpoint with multi RF–multi CP [CP_{n-1}, CP_n] showing the best results compared with the previous two models tested. This suggests that to better

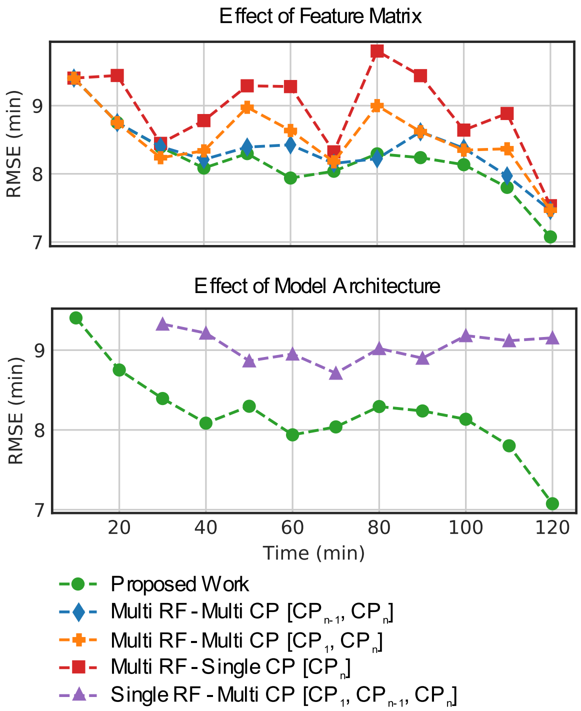


Fig. 5. TTC RMSE while varying: 1) the input feature matrix or 2) the number of models trained to predict TTC. For 1), feature matrices were altered from the proposed method by removing baseline features (multi-RF–multi-CP [CP$_1$, CP$_n$]), the previous checkpoint features (multi-RF–multi-CP [CP$_{n-1}$, CP$_n$]) or both the previous checkpoint and baseline features (multi-RF–single CP). For 2), the number of models was altered by training a single model to predict TTC (single RF–multi-CP) over the entire march. The proposed method indicates the benefit of multiple checkpoint and models for TTC estimation.

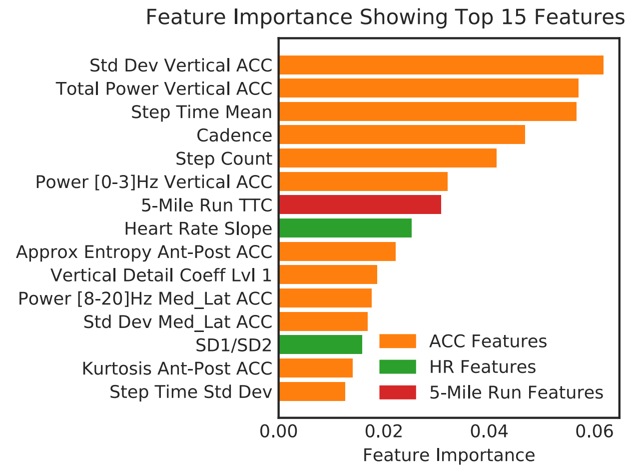


Fig. 6. Top 15 features of the proposed RF regression model. Vertical ACC standard deviation was the most important feature along with other gait-related features. SCT and CT features did not have a high enough importance to be included in this list.

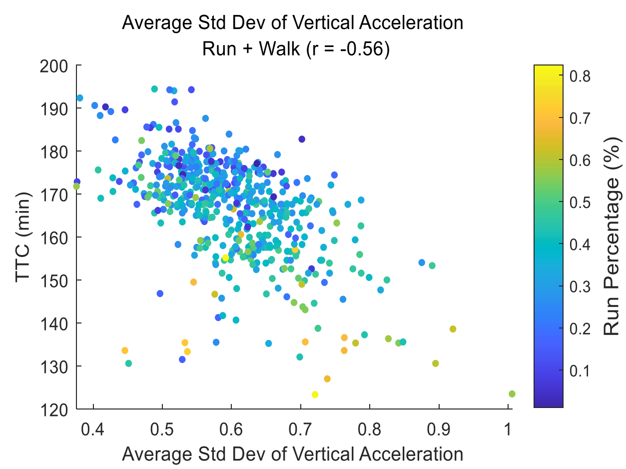


Fig. 7. Relationship of vertical ACC standard deviation and TTC-color coded by percentage of run windows.

estimate TTC at a point in the march requires not only current physiological and movement data but also the past data.

Using the single RF–multi-CP configuration maintains consistent performance throughout the march at a higher RMSE than the proposed approach. The increase in RMSE suggests that different environmental factors throughout the march may have an effect on TTC estimation. For example, high strain periods in physiological measurements may be due to general fatigue that affects lower performing soldiers or large changes in inclines that affect all participating soldiers. As such, dividing the march into distinct checkpoints implicitly provides an environmental context that is common between all soldiers tested.

C. Feature Importance

Fig. 6 shows the top 15 features across all trained models. The most important feature was the average standard deviation of vertical ACC and the average power of vertical ACC. Both features indicate a sense of force applied in the vertical direction. In addition, many of the top features were gait-related, such as step time, cadence, and step count, which are the important parameters for calculating distance and speed.

Fig. 7 shows further investigation into the relationship between high-vertical ACC standard deviation and TTC, color coded by the percentage of windows that the subject was spent running. In this case, the running and walking boundaries are defined on a subject-by-subject case using cadence as a discriminatory feature [27]. We can see a negative correlation between TTC and vertical ACC standard deviation. In addition, subjects with a high variance of vertical ACC are usually associated with a high percentage of running windows. This indicates that increased vertical ACC is closely related to high cadence, both of which contribute to a lower TTC.

We also observe that the 5-mile TTC was important to predicting 12-mile TTC. This suggests that prior exercise information can be useful for predicting TTC in another exercise.

HR slope was important in predicting TTC. HR and HRV features have been shown to be indicators of energy expenditure [28]. However, there are fewer high-contributing HR features compared with ACC features, indicating that ACC features are more important for predicting TTC.

Further of interest is that although ECT and SKT were included as features for TTC estimation based on the fact that they serve as important markers for fitness in the context

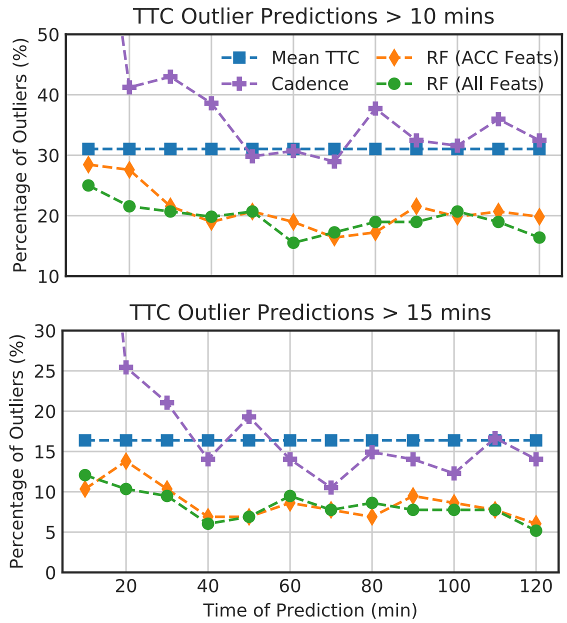


Fig. 8. Percentage of TTC outlier predictions using the absolute error between the estimated and true TTC as a criterion. For better visualization, percentage plot ranges exclude the 10-min outlier percentage value when using the cadence model (90.35% for 10-min threshold and 74.56% for the 15-min threshold).

of heat exertion, it was found that these features do not significantly affect TTC prediction and do not factor in the top 15 contributing features of our models.

D. Outlier Analysis

To analyze the model performance, we examined the number of TTC predictions that were severely skewed from the true TTC values. Fig. 8 shows the percentage of subjects at each checkpoint, whose absolute TTC prediction error exceeded a threshold of 10 or 15 min. A 10 min was chosen as a threshold to match the march segmentation duration, while 15 min was chosen to contain 1 standard deviation of the overall TTC distribution (166.86 $\pm$ 11.89 min). Both models have fewer outliers compared to using the mean TTC method. There is also a slight downward trend of outliers as the exercise progresses, indicating the model becomes more accurate over time. While the outlier percentage using a threshold of 10 min drops from 25% to 15% with the proposed model, the outlier percentage using 15 min drops from 11% to 5%. This discrepancy of outliers between the two thresholds suggests that a large portion of outliers are within the error range of 10–15 min. These outliers correspond to subjects with a TTC in the higher or lower range, as shown in Fig. 4. We suspect that the model estimates TTC poorer on these subjects than those with moderate TTC because of the low number of available subjects in the TTC upper and lower ranges. More data with TTC labels at the extremities may be needed to enable the model to learn behaviors at these ranges.

IV. Limitations

A limitation of the proposed work is the requirement of training multiple models to estimate TTC sequentially. While random forests are lightweight and the feature extraction done in this work is computationally simple, future work should investigate the deployment capabilities of the proposed framework in a portable system. Specifically, the data size of the random forest and the computational power effect on battery life should be examined.

Because signal quality was done rudimentarily on a global scale for each subject, intermittent sensor noise may negatively affect estimation performance. For ACC in particular, aside from egregious data, the distinction between abnormal ACC data and normal ACC data is not always clear [29]. Confounding factors such as shifts in sensor placement and posture may create changes in measurements that are not easily observable. This becomes even more difficult given the noisy measurement environment during the march. The addition of a more robust signal quality block may prevent corrupted data from producing misleading results.

Another apparent limitation is that the study was limited to 12-mile ruck march data collected at Fort Benning. Different terrains or exercises may have different sensor measurement patterns. For example, exercises with high intensity portions in the beginning may have poor estimation performance when using a model trained on exercises with low-intensity portions in the beginning. For this application, the intended use is for longitudinal studies using the same course. Future work includes TTC estimation using additional global fitness features (e.g., physical training results) and different march routes to determine the model's generalizability. In addition, activity recognition models can be run in parallel to give more context to ACC and physiological measurements.

V. Conclusion

In this work, we demonstrated a TTC prediction framework for structured marches using simple measurements of ACC, heart rate, and SKT from portable wearable devices. We achieved an average absolute error of 5.23 min by 2/3 of the expected completion time with few outlier predictions. The use of separate models for each checkpoint takes advantage of the march's structured nature to reduce the prediction error as the march progresses. The proposed model can be trained on an incoming group of soldiers and can be used for longitudinal tracking of TTC as they go through these structured marches multiple times during basic training. Such performance metrics are important for military personnel to evaluate the physical capabilities and readiness of soldiers before field deployment. While this work focused on military ruck marches, the proposed methodology can be extended to other structured athletics disciplines, such as cross-country and track-and-field. For these situations and when applying the proposed methodology to other disciplines, new training datasets representative of the exercise being performed would be necessary to achieve the desired results.

Finally, this work primarily discusses the development of a performance estimation algorithm for structured physical exercises and not its exact implementation in an edge or cloud computing architecture with real-time data streaming capabilities. However, Georgia Tech Research Institute (GTRI) is currently collaborating with the U.S. Army in the development of such a system, wherein live data can wirelessly stream from physiological sensors to a cloud server for real-time geolocation, postprocessing, and visualization. In the future, we envision integrating the TTC estimation model presented in this work into this infrastructure.

ACKNOWLEDGMENT

David J. Lin, Kristine L. S. Richardson, Cem O. Yaldiz, and Omer T. Inan are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: dlin91@gatech.edu; kscott43@gatech.edu; cyaldiz3@gatech.edu; omer.inan@ece.gatech.edu).

Aprameya Satish and Alessio Medda are with the Georgia Tech Research Institute, Atlanta, GA 30332 USA (e-mail: aprameya.satish@gtri.gatech.edu; alessio.medda@gtri.gatech.edu).

Sungtae An is with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: sungtae.an@cc.gatech.edu).

Mark Buller, Kyla Driver, Emma Atkinson, and Timothy Mesite are with the U.S. Army Research Institute of Environmental Medicine, Natick, MA 01516 USA (e-mail: mark.j.buller.civ@health.mil; kyla.a.driver.civ@health.mil; emma.y.atkinson.ctr@health.mil; timothy.j.mesite.ctr@health.mil).

Christopher King is with the Combat Capabilities Development Command (DEVCOM), Maneuver Center of Excellence, Fort Moore, GA 31905 USA (e-mail: christopher.j.king22.civ@health.mil).

REFERENCES

[1] J. J. Knapik, W. Rieger, F. Palkoska, S. V. Camp, and S. Darakjy, "United States army physical readiness training: Rationale and evaluation of the physical training doctrine," *J. Strength Conditioning Res.*, vol. 23, no. 4, pp. 1353–1362, Jul. 2009, doi: 10.1519/JSC.0b013e318194df72.

[2] T. Wyss, J. Scheffler, and U. Mäder, "Ambulatory physical activity in Swiss army recruits," *Int. J. Sports Med.*, vol. 33, no. 9, pp. 716–722, Sep. 2012, doi: 10.1055/s-0031-1295445.

[3] M. Cardinale and M. C. Varley, "Wearable training-monitoring technology: Applications, challenges, and opportunities," *Int. J. Sports Physiol. Perform.*, vol. 12, no. s2, pp. 55–62, Apr. 2017, doi: 10.1123/ijspp.2016-0423.

[4] W. J. Tharion and R. W. Hoyt, "Form factor evaluation of open body area network (OBAN) physiological status monitoring (PSM) system prototype designs," U.S. Army Res. Inst. Environ. Med., Natick, MA, USA, Tech. Rep. T18-06, May 2018. Accessed: Jun. 29, 2022.

[5] A. Ometov et al., "A survey on wearable technology: History, state-of-the-art and current challenges," *Comput. Netw.*, vol. 193, Jul. 2021, Art. no. 108074, doi: 10.1016/j.comnet.2021.108074.

[6] S. Seneviratne et al., "A survey of wearable devices and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2573–2620, 4th Quart., 2017, doi: 10.1109/COMST.2017.2731979.

[7] E. Abdi, H. S. Mariv, A. Deljouei, and H. Sohrabi, "Accuracy and precision of consumer-grade GPS positioning in an urban green space environment," *Forest Sci. Technol.*, vol. 10, no. 3, pp. 141–147, Jul. 2014, doi: 10.1080/21580103.2014.887041.

[8] V. A. Paz-Soldan et al., "Strengths and weaknesses of global positioning system (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: Findings from Iquitos, Peru," *PLoS Neglected Tropical Diseases*, vol. 8, no. 6, p. e2888, Jun. 2014, doi: 10.1371/journal.pntd.0002888.

[9] H. Xing, J. Li, B. Hou, Y. Zhang, and M. Guo, "Pedestrian stride length estimation from IMU measurements and ANN based algorithm," *J. Sensors*, vol. 2017, pp. 1–10, Feb. 2017, doi: 10.1155/2017/6091261.

[10] A. Köse, A. Cereatti, and U. D. Croce, "Bilateral step length estimation using a single inertial measurement unit attached to the pelvis," *J. NeuroEngineering Rehabil.*, vol. 9, no. 1, p. 9, Feb. 2012, doi: 10.1186/1743-0003-9-9.

[11] Y. Mao, T. Ogata, H. Ora, N. Tanaka, and Y. Miyake, "Estimation of stride-by-stride spatial gait parameters using inertial measurement unit attached to the shank with inverted pendulum model," *Sci. Rep.*, vol. 11, no. 1, p. 1, Jan. 2021, doi: 10.1038/s41598-021-81009-w.

[12] S. E. Crouter, J. R. Churilla, and D. R. Bassett, "Estimating energy expenditure using accelerometers," *Eur. J. Appl. Physiol.*, vol. 98, no. 6, pp. 601–612, Dec. 2006, doi: 10.1007/s00421-006-0307-5.

[13] M. J. Buller et al., "Estimation of human core temperature from sequential heart rate observations," *Physiological Meas.*, vol. 34, no. 7, pp. 781–798, Jun. 2013, doi: 10.1088/0967-3334/34/7/781.

[14] C. A. Clermont, L. C. Benson, W. B. Edwards, B. A. Hettinga, and R. Ferber, "New considerations for wearable technology data: Changes in running biomechanics during a Marathon," *J. Appl. Biomechanics*, vol. 35, no. 6, pp. 401–409, Dec. 2019, doi: 10.1123/jab.2018-0453.

[15] D. Herzig, B. Asatryan, N. Brugger, P. Eser, and M. Wilhelm, "The association between endurance training and heart rate variability: The confounding role of heart rate," *Frontiers Physiol.*, vol. 9, p. 756, Jun. 2018, Accessed: May 27, 2022.

[16] A. L. Baggish and M. J. Wood, "Athlete's heart and cardiovascular care of the athlete," *Circulation*, vol. 123, no. 23, pp. 2723–2735, Jun. 2011, doi: 10.1161/CIRCULATIONAHA.110.981571.

[17] B. A. Telfer, K. Byrd, and P. P. Collins, "Open body area network physiological status monitor," *Lincoln Lab. J.*, vol. 24, no. 1, pp. 1–17, 2020.

[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[19] U. Grömping, "Variable importance assessment in regression: Linear regression versus random forest," *Amer. Statistician*, vol. 63, no. 4, pp. 308–319, Nov. 2009, doi: 10.1198/tast.2009.08199.

[20] E. Fortune, V. Lugade, M. Morrow, and K. Kaufman, "Validity of using tri-axial accelerometers to measure human movement—Part II: Step counts at a wide range of gait velocities," *Med. Eng. Phys.*, vol. 36, no. 6, pp. 659–669, Jun. 2014, doi: 10.1016/j.medengphy.2014.02.006.

[21] M. Costa, C.-K. Peng, A. L. Goldberger, and J. M. Hausdorff, "Multiscale entropy analysis of human gait dynamics," *Phys. A, Stat. Mech. Appl.*, vol. 330, nos. 1–2, pp. 53–60, Dec. 2003, doi: 10.1016/j.physa.2003.08.022.

[22] R. Moe-Nilssen and J. L. Helbostad, "Estimation of gait cycle characteristics by trunk accelerometry," *J. Biomech.*, vol. 37, no. 1, pp. 121–126, Jan. 2004, doi: 10.1016/S0021-9290(03)00233-1.

[23] M. N. Sawka and A. J. Young, "Physiological systems and their responses to conditions of heat and cold," Army Res. Inst. Environ. Med., Thermal and Mountain Medicine Division, Natick MA, USA, Tech. Rep. MISC 03-12, Jan. 2006. Accessed: Jun. 12, 2022.

[24] G. L. Pierpont, D. R. Stolpman, and C. C. Gornick, "Heart rate recovery post-exercise as an index of parasympathetic activity," *J. Autonomic Nervous Syst.*, vol. 80, no. 3, pp. 169–174, May 2000, doi: 10.1016/S0165-1838(00)00090-4.

[25] G. L. Pierpont and E. J. Voth, "Assessing autonomic function by analysis of heart rate recovery from exercise in healthy subjects," *Amer. J. Cardiology*, vol. 94, no. 1, pp. 64–68, Jul. 2004, doi: 10.1016/j.amjcard.2004.03.032.

[26] P. E. Martin and R. C. Nelson, "The effect of carried loads on the walking patterns of men and women," *Ergonomics*, vol. 29, no. 10, pp. 1191–1202, Oct. 1986, doi: 10.1080/00140138608967234.

[27] C. Chase, "Cadence as an indicator of the walk-to-run transition," M.S. thesis, Degree Program Kinesiology, Univ. Massachusetts Amherst, Amherst, MA, USA, 2020, doi: 10.7275/16698528.

[28] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, p. 258, Sep. 2017.

[29] C. Rich, M. Geraci, L. Griffiths, F. Sera, C. Dezateux, and M. Cortina-Borja, "Quality control methods in accelerometer data processing: Identifying extreme counts," *PLoS ONE*, vol. 9, no. 1, Jan. 2014, Art. no. e85134.

David J. Lin (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2019, and the M.S. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2022, where he is currently pursuing the Ph.D. degree in electrical and computer engineering.

In 2019, he joined the Inan Research Lab, Georgia Institute of Technology, as a Graduate Research Assistant. His research focuses on signal processing and machine learning for cardiovascular physiological signals with an emphasis on real-time monitoring during daily activities.

Aprameya Satish (Member, IEEE) received the M.S. degree in electrical engineering and the Ph.D. degree in mechanical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2020, with a focus on underwater acoustics, signal processing, and applied machine learning.

Upon graduation, he joined the Georgia Tech Research Institute (GTRI), Atlanta, as a Research Engineer II, where he has worked on and led signal processing, machine learning, and field-testing efforts in programs related to physiological health monitoring in military populations, acoustic and infrasound beamforming, synthetic aperture radar, and aircraft predictive maintenance.

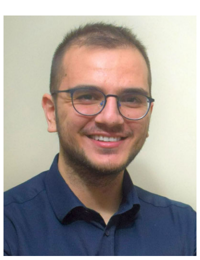
Kristine L. S. Richardson (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in electrical and computer engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2018 and 2020, respectively, where she is currently pursuing the Ph.D. degree in electrical and computer engineering.

Throughout her undergraduate degree, she interned at OFS Fitel LLC, Norcross, GA, USA, where she researched optical fiber data transmission, and also at the Georgia Tech Research Institute, Atlanta, where she researched in the field of biomedical engineering. In 2019, she joined the Inan Research Lab, Georgia Tech, as a Graduate Research Assistant. Her research focuses on signal processing and machine learning for biomedical signals to assess joint health.

Dr. Richardson received the Szlam Fellowship and the President's Fellowship from Georgia Tech, in 2019 and 2020, respectively.

Sungtae An received the B.S. degree in electrical and computer engineering from Hanyang University, Seoul, South Korea, in 2010, and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2012. He is currently pursuing the Ph.D. degree in computer science with the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA.

In 2020, he joined the Inan Research Lab, Georgia Tech, as a Graduate Research Assistant. His research focuses on deep learning with less supervised environments, including weakly supervised, semisupervised, and transfer learning, particularly for human activity recognition and physiological signal analysis.



Cem O. Yaldiz received the B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2021. He is currently pursuing the Ph.D. degree in robotics with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

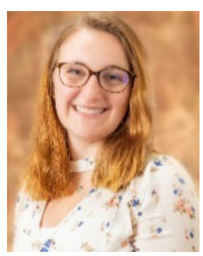
He is conducting research at the Inan Research Lab, Georgia Institute of Technology. His research interests include applied machine learning, data-driven decision-making, representation learning, and multimodal information fusion.

Mark Buller received the Ph.D. degree in computer science from Brown University, Providence, RI, USA, in 2015, with a focus on computational physiology.

He is a Data Scientist and the Advanced Technology Team Leader with the Thermal and Mountain Medicine Division, U.S. Army Research Institute of Environmental Medicine, Natick, MA, USA. He has over 20 years of experience in designing and fielding ambulatory monitoring systems for warfighters. He has authored more than 100 publications. His current research interests include real-time algorithms that determine health state from wearable sensors, and performance optimization from physiological feedback. Dr. Buller is currently serving as the Chair for a Real-Time Monitoring NATO Working Group.

Kyla Driver received the B.S. degree in biomedical engineering and the M.S. degree in science, technology, and public policy from the Rochester Institute of Technology, Rochester, NY, USA, in 2019.

Throughout her undergraduate degree, she interned at the University of Vermont Medical Center, Burlington, VT, USA, as a Clinical Engineer, developing cyber security plans for protected health information systems; University of Rochester, Rochester, as a National Science Foundation Research Fellow, testing drug effectiveness for presbyopia reversal; and Philips Healthcare, Andover, MA, USA, as an Research and Development Engineer Intern, completing product evaluation on a novel technology for arterial plaque ablation. In 2019, she joined the U.S. Army Research Institute of Environmental Medicine, Natick, MA, USA, as an Oak Ridge Institute for Science and Education Fellow and transitioned to the Department of Defense (DoD) civilian in 2020. Throughout her time at USARIEM, she has worked as a Biomedical Engineer and a Research Study Coordinator. Her research focuses on conducting data collection for algorithm development and providing solutions to the warfighters to help with risk mitigation of heat-related injuries in military training.



Emma Atkinson received the B.S. degree in biomedical engineering with a concentration in bioinstrumentation from the University of Connecticut, Storrs, CT, USA, in 2019.

In 2019, she joined the U.S. Army Research Institute of Environmental Medicine, Natick, MA, USA, as an Oak Ridge Institute for Science and Education Research Fellow and transitioned to a Department of Defense (DoD) civilian position in 2023. During her time at USARIEM, she has worked as a biomedical engineer conducting research on how the feedback of real-time physiologic data during rigorous military training events can predict the onset of exertional heat illness. Her research interests include physiological signal analysis, algorithm development, and the use of real-time physiological status monitors during military training events to prevent the occurrence of heat injuries.

Timothy Mesite received the B.S. degree in computer science from Worcester State University, Worcester, MA, USA, in 2019.

He began working at the U.S. Army Research Institute of Environmental Medicine, Natick, MA, USA, as an Oak Ridge Institute for Science and Education Fellow, supporting the Thermal Mountain Medicine Division with creation, maintenance, and troubleshooting of research devices both commercial and made in-house.

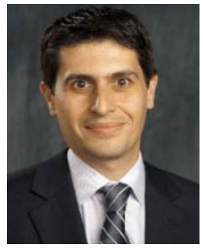
Christopher King received the B.S. degree in liberal studies from the Excelsior College, Albany, NY, USA, in 2004.

He was with United States Army Special Forces, from 1982 to 2004. In 2004, he became a member of the Battle Lab Integration Team, Natick Soldier Research, Development, and Engineering Center, and in 2009, he joined NSRDEC, Natick, MA, USA, as a member of the Operational Force Interface Group. He currently serves as the Deputy NSRDEC Technical Representative of the Maneuver Center of Excellence (MCoE), Fort Moore, GA, USA, where he coordinates MCoE activities across several U.S. Army regiments.

Omer T. Inan (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 2004, 2005, and 2009, respectively.

He joined ALZA Corporation, Mountain View, CA, USA, (A Johnson and Johnson Company), in 2006, where he designed micropower circuits for iontophoretic drug delivery. In 2007, he joined Countryman Associates Inc., Menlo Park, CA, USA, where he was the Chief Engineer, involved in designing and developing high-end professional audio circuits and systems. From 2009 to 2013, he was a Visiting Scholar with the Department of Electrical Engineering, Stanford University. From 2013 to 2018, he was an Assistant Professor of Electrical and Computer Engineering with the Georgia Institute of Technology, Atlanta, GA, USA, where he is currently the Linda J. and Mark C. Smith Chair in Bioscience and Bioengineering, an Associate Professor of Electrical and Computer Engineering, and an Adjunct Associate Professor of Biomedical Engineering. He has authored or coauthored more than 230 technical articles in peer-reviewed international journals and papers in conferences and has eight issued patents and more than 350 technical articles in peer-reviewed international journals and conferences and has 15 issued patents. His research focuses on noninvasive physiologic sensing and modulation for human health and performance, including for chronic disease management, acute musculoskeletal injuries and disorders, and pediatric care.

Dr. Inan is a Fellow of the American Institute for Medical and Biological Engineering (AIMBE) and the American College of Cardiology (ACC). He received the Gerald J. Lieberman Fellowship in 2009, the Lockheed Dean's Excellence in Teaching Award in 2016, the Sigma Xi Young Faculty Award in 2017, the IEEE Sensors Early Career Award in 2018, the Office of Naval Research Young Investigator Award in 2018, the National Science Foundation CAREER Award in 2018, and the Academy Award for Technical Achievement from the Academy of Motion Picture Arts and Sciences (The Oscars) in 2021. He is an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING and the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, a Theme 10 Co-Editor of the IEEE Engineering in Medicine and Biology Conference, an Associate Editor of the IEEE Biomedical and Health Informatics Conference, a Vice Chair of the IEEE Technical Committee (TC) on Wearable Biomedical Sensing Systems, an Invited Member of the IEEE TC on Translational Engineering for Healthcare Innovation and the TC on Cardiopulmonary Systems, and a technical program committee member or a track chair for several other major international biomedical engineering conferences. He was a National Collegiate Athletic Association (NCAA) All-American in the discus throw for three consecutive years (2001–2003).



Alessio Medda (Senior Member, IEEE) received the M.S. degree in electronic engineering from the University of Cagliari, Cagliari, Italy, in 1993, and the Ph.D. degree in electric and computer engineering from Florida State University, Tallahassee, FL, USA, in 2009.

From 2009 to 2010, he was a Visiting Assistant Professor with the Department of Electrical and Computer Engineering, Florida State University. In 2010, he joined the U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL, USA, as a Signal Analyst, working on algorithm development and data analysis for the Helmet Sensor Project. In 2011, he joined the Georgia Tech Research Institute (GTRI), Atlanta, GA, USA, as a Research Engineer II, he was promoted to a Senior Research Engineer in 2015 and to a Principal Research Engineer, in 2022. In 2019, he became the Head of the Biosensing and Signal Analytics Branch, Aerospace, Transportation, and Advanced System Laboratory, GTRI. He has authored more than 50 publications in peer-reviewed journals and conferences papers, and he has two issued patents. His research interests are in physiological and environmental sensing for civilian and military applications, acoustics beamforming, application of wavelets, and machine learning.

Dr. Medda was a recipient of the 2013 GTRI Junior Research Excellence Award; the 2013, 2015, and 2016 Outstanding Technical Achievement Award; and the two Georgia Tech Teaching Fellowship Awards. He was the Chair of the IEEE Atlanta Signal Processing Chapter from 2014 to 2020. He is the Chair of the IEEE Atlanta Section and the future General Chair of the SoutheastCon 2024 conference. He served as an Assistant Editor for *IEEE Signal Processing Magazine* (eNewsletter).