Estimating Energy Expenditure with Multiple Models using Different Wearable Sensors

Božidara Cvetković, Radoje Milić, and Mitja Luštrek

***Abstract*—This paper presents an approach to designing a method for the estimation of human energy expenditure (EE). The approach first evaluates different sensors and their combinations. After that, multiple regression models are trained utilising data from different sensors. The EE-estimation method designed in this way was evaluated on a dataset containing a wide range of activities. It was compared against three competing state-of-the-art approaches, including the BodyMedia Fit armband, the leading consumer EE estimation device. The results show that the proposed method outperforms the competition by up to 10.2 percentage points.**

***Index Terms*—Pervasive computing, Sensor fusion, Wearable sensors, Machine learning**

I. Introduction

IT is widely accepted that sufficient physical activity has a positive impact on health and well-being, and that inactivity contributes to increased risk of development of various diseases [1][2][3]. The reasons for inactivity include sedentary work, lack of time and lack of motivation for physical activity during leisure time [4]. Pervasive technology is well-suited to motivating people to be physically active, and it can also estimate the amount of activity performed during a day, which provides a starting point for the motivation. Furthermore, estimating the amount of activity can help users more accurately control their diet, which is significant both for healthy people (e.g., weight control, athletes) and individuals who suffer from diseases whose management requires paying close attention to the diet (e.g., diabetes).

The amount of physical activity (PA) can be expressed as the energy expenditure (EE). It is usually measured in metabolic equivalents of task (MET), where 1 MET corresponds to the energy expended at rest. MET values range from 0.9 during sleeping to over 20 during extreme exertion. PA of up to 3 MET is considered light exertion, from 3 to 6 MET moderate exertion and above 6 MET vigorous exertion.

The methods that reliably estimate the EE are expensive, cumbersome and sometimes limited to EE estimation over longer periods of time [5]. Direct calorimetry, which measures the heat produced by the human body, is the most accurate method. Its downside is that it can be used only in a controlled environment such as a laboratory since it requires the confinement of the subject to a chamber. Indirect calorimetry measures the carbon dioxide production and oxygen consumption. This method can be used outside the laboratory, but is still not appropriate for free-living conditions since it requires a breathing mask. Doubly labelled water measures the amount of exhaled carbon dioxide by tracking its amount in ingested water labelled by deuterium and oxygen-18. This method can be used in free-living conditions, but it measures and averages the EE over at least 24-hour period by sampling body fluids (saliva, urine or blood), which are afterwards processed in a laboratory to determine the EE from isotope concentration. Since these methods are not suitable for pervasive monitoring of PA, the question is how well can their accuracy be matched by more convenient and inexpensive wearable sensors.

This paper presents an approach to designing a method for the estimation of human energy expenditure (EE) with wearable sensors by utilising machine-learning techniques and sensor fusion. It presents a systematic evaluation of eight sensors and their combinations for the estimation of EE. The evaluation serves to design a new method for the estimation of EE. The evaluation of this method were done against the BodyMedia Fit armband [6], a state-of-the-art consumer device for EE estimation, and against the recent research done by Altini et al. [7] and by Gjoreski et al [8]. The proposed method outperformed all three competing approaches. The validation of the method was done on a separate dataset.

II. Related Work

The early attempts to estimate human EE using pervasive technology used a single accelerometer attached to the user's body, and a single linear regression model. Since this approach proved insufficiently accurate for the EE estimation of light and vigorous PA, Crouter et al. [9] employed different regression models for different activities. Their system first recognises the type of the activity (sedentary, ambulatory or lifestyle). If the recognised type is sedentary, the estimated EE equals 1 MET, and in the case of ambulatory or lifestyle activity the respective linear regression models are used. The shortcomings of this approach are the underestimation of sedentary activities and the relative simplicity of the activity

Date.

B. Cvetković is with the Jožef Stefan Institute, Department of Intelligent Systems, Jamova cesta 39, 1000 Ljubljana, Slovenia. She is also with Jožef Stefan Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia. (e-mail: boza.cvetkovic@ijs.si).

R. Milić is with the Faculty of Sports, University of Ljubljana, Ljubljana, Slovenia (e-mail: radoje.milic@fsp.uni-lj.si).

M. Luštrek is with the Jožef Stefan Institute, Department of Intelligent Systems, Jamova cesta 39, 1000 Ljubljana, Slovenia (e-mail: mitja.lustrek@ijs.si).

recognition and regression methods compared to later work [10].

Further enhancements of EE estimation include improved activity recognition [11] and the employment of additional sensors, such as a heart-rate monitor, which improves the recognition of the PA intensity. Lin et al. [12] used three accelerometers (wrist, waist, ankle) and an ECG sensor (measuring the heart rate) to recognise activities with similar intensities (three types) and estimate the EE using three artificial neural networks, one per intensity type. Altini et al. [13] also combined an ECG sensor and an accelerometer. Their activity-recognition procedure clusters the activities into more fine-grained groups than in the previous studies (lying, sitting, standing, high whole-body motion (HWBM), walking, running and cycling), for each of which they use an activity-specific regression model to estimate the EE. In a more recent study, Altini et al. [7] showed that two accelerometers (chest and ankle) are enough for accurate activity recognition and EE estimation. For the EE estimation, they proposed a hybrid approach. For lying, sitting and standing with low body motion, the EE is estimated with a MET lookup method, which returns a static value for each person and activity. The value is computed with a regression model trained on anthropometric features and static MET values. For HWBM, walking, running and cycling, activity-specific regression models utilising sensor data are used.

Recent studies analyse even more sensors to be utilised for EE estimation. Gjoreski et al. [8] utilise multi-sensor fusion with eight physical sensors (four temperature sensors, a breath-rate sensor, two accelerometers and a heart-rate monitor), and two virtual sensors (activity and accelerometer peak count). Multiple models are built not only for different activities, but also for different intervals of values of other (physical and virtual) sensors. A similar approach is used by Vyas et al. [14], combining data from accelerometer and temperature-related physiological sensors embeded into an armband. Each sensor represents one or more contexts upon which context-specific regression model are used and the estimations among contexts are averaged for the final estimation. This approach is implemented the BodyMedia Fit [6] consumer device. Lee et al. [15] compared several single- and multi-sensor devices in free-living conditions. The results of the experiments clearly favoured the BodyMedia Fit.

The approach proposed in this paper systematically analyses and evaluates the number and contribution of the various sensors and features computed from their outputs and combination of models, aiming to design an accurate method for estimation of EE. The designed method is composed of set of regression models which are used according to the recognised activity, and differ in the number and type of utilized sensors. Experimental results proved the value of such analysis by showing that the designed method significantly outperforms the methods by Altini et al. [7] and Gjoreski et al. [8], and the BodyMedia FIT armband.

III. System Description

The pipeline of the system we used for the estimation of EE is shown in Fig. 1. The core of the system consist of three consecutive modules: the data synchronisation, the activity recognition and the EE estimation module. The input to the system are raw data from an array of sensors which differ by their modality and frequency. The output is the estimated EE.

The sensor data are retrieved from the following devices: two Shimmer accelerometers [16], one on the chest and one on the thigh, the Zephyr BioHarness chest strap [17], the BodyMedia Fit armband [6] and the Cosmed K4b^2 indirect calorimeter [18]. The device models, sensors per device and their frequencies are presented in Table I. The Cosmed K4b^2 indirect calorimeter provides the reference energy expenditure used to train the regression models. The BodyMedia device is used for two purposes: first, the raw measurements of its sensors are used as features in the regression models, and second, its EE estimation is used for the final comparison with the proposed approach.

The data synchronisation is performed as follows. Composite sensor readings termed snapshots are constructed at 33 Hz (to get fresh readings from both accelerometers for each snapshot). Each snapshot contains data that represent the last received reading from all the connected sensors (excluding the Cosmed K4b^2 indirect calorimeter). The first snapshot is constructed when at least one measurement from each connected sensor is received. Afterwards, one snapshot is constructed every 30 ms regardless of whether all the connected sensors have sent new measurements.

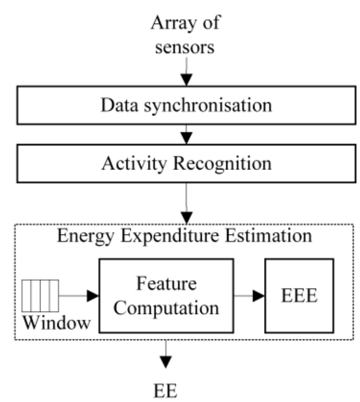

Fig. 1. System pipeline for energy expenditure estimation.

TABLE 1
Devices, sensors and their properties

Device	Sensor	Frequency
Shimmer2	Accelerometer (ACC)	50 Hz
Zephyr BioHarness™	Heart rate (HR) Breath rate (BR) Skin temperature (ST)	1 Hz
BodyMedia Fit (SenseWear Software 8.0)	Near-body temperature (NBT) Galvanic skin response (GSR) Skin temperature (ST) Estimated EE	1 per min
Cosmed K4b^2	Measured EE	20 per min

The activity recognition used in this research was adopted from Kozina et al. [19] and is briefly described here. Snapshots with raw accelerometer data are collected into 2-second windows for which, a number of activity-recognition features are computed. These features constitute the feature vector fed into an activity-recognition classifier trained with the Random Forest algorithm [20], The accuracy of the classifier is 90% with an accelerometer on the chest and 92% with accelerometers on the chest and thigh.

The snapshots completed with the recognised activity are queued in the EE estimation module where they are collected into 10-second windows, each overlapping with previous one by half of its length. Like in the activity recognition, for each window a number of features are computed. The computed features constitute the feature vector that is fed into the appropriate EE-estimation regression model, depending on the recognised activity. The output is the EE expressed in MET.

The paper is focused on the last module of the system pipeline, whose construction is explained in detail in Section V. It is constructed in multiple steps, each of which depends on the experimental results from the previous step, so the experimental results are also presented in Section V, proceeded by the experimental setup in Section IV.

IV. Experimental Setup

We collected two datasets: one for the evaluation (*E*) of different combinations of sensors and regression models to produce the final EE estimation method, and one only for the validation (*V*) of the method. The *E* dataset contains the data of ten healthy volunteers (eight male and two female, age from 24 to 33, body mass index from 20 to 28.9) performing a predefined set of activities composing scenarios. The *V* dataset contains the data of five different healthy volunteers (three male and two female, age from 25 to 30, body mass index from 20.5 to 24.8) performing slightly different activities and scenarios. Both datasets are presented in Table II. Note that the *V* dataset was recorded 14 months after the *E* dataset. Each volunteer was equipped with sensors presented in Table I. All volunteers refrained from eating and drinking (except for water) in the 12 hours before the experiment.

The scenarios were drawn up to include both everyday activities, such as resting, cleaning, shovelling and office work, and exercise, such as normal and fast walking, running and normal and vigorous stationary cycling. Moreover, the activities were ordered by increasing EE, the light resting in the beginning building up to the vigorous running at the end. This ensured that the body processes stimulated by vigorous activities did not distort the data collected during less vigorous activities. For the same reason five-minute rests were imposed between the scenarios containing moderate and vigorous activities. The activities were performed in a sports laboratory equipped with a treadmill, a stationary bicycle, office furniture, a bed and an area used as a kitchen. The walking and running activities were controlled by the speed and inclination of the treadmill, and the stationary cycling was controlled by the power in Watt (W). The reference EE measured by the Cosmed indirect calorimeter ranged from 0.9 MET to 12 MET. We collected approximately 1 hour and 45 minutes of data or approximately 410,000 measurements per volunteer.

TABLE II
Scenarios performed by volunteers. The dataset columns show which activities were present in which dataset (evaluation E, validation V) and the last column represents the average EE measured by the Cosmed indirect calorimeter.

Scenario	Activities	Duration (min)	Dataset E	Dataset V	Avg. EE (MET)
Resting	Lying	15	✓	✓	1.2
Basic postures	Sitting	30	✓	✓	1.15
	Standing		✓	✓	1.21
	Walking		✓	✓	1.37
	Transition		✓	✓	1.97
	All fours		✓	✓	2.22
	Kneeling		✓	✓	1.45
	Leaning		✓	✓	1.85
Office work, typing	Sitting	6	✓	✓	1.17
Lying exercising, stretching	Lying	6	✓	✓	2.12
Kitchen chores	Standing	6	✓	✓	1.68
	Walking		✓	✓	2.02
Scrubbing the floor	Kneeling	6	✓		3.20
	All fours		✓		3.03
Moping the floor	Standing	6		✓	1.78
	Walking			✓	2.01
Shoveling snow	Standing	6	✓		3.06
	Walking		✓		3.60
Walking	Walking 4km/h	6	✓	✓	3.02
	Walking 6 km/h	6	✓	✓	4.54
	Walking 4km/h*	6		✓	6.53
	Walking			✓	5.30
	Transition	3		✓	/
	Sitting			✓	2.89
Walking carrying burden	Walking 4km/h**	6		✓	3.44
	Walking			✓	2.71
	Transition	3		✓	/
	Sitting			✓	1.77
Stationary cycling	Cycling 1W	6	✓	✓	4.22
	Cycling 2W	6	✓	✓	6.30
	Walking			✓	5.46
	Transition	3		✓	/
	Sitting			✓	2.62
Running	Running 8km/h	6	✓	✓	7.70
	Walking			✓	7.55
	Transition	3		✓	/
	Sitting			✓	3.39

*10% inclination, ** Female 3 kg, male 6 kg burden

V. Development, Evaluation and Validation of the EE Estimation Method

When developing the EE estimation method, the goal was to systematically evaluate the sensors available in our recordings and to choose the regression models that ensure the most accurate estimation. We used the *E* dataset for this as well as for the evaluation of the final method against the competition. The *V* dataset was used to validate our method on different data to guard against overfitting to the training data.

A. Evaluation of sensors

The starting point was to evaluate each sensor from the sensor array to decide on the initial sensor configuration. This initial configuration was afterwards enriched with additional sensors from the sensor array.

Most sensors provided only one feature to be used in regression models. One exception are the skin-temperature (ST) and near-body-temperature (NBT) sensors, which provided an additional feature: deltaT = NBT – ST. The other exception are accelerometers, which provide over a hundred features [21]. Because of that we used feature selection returning 30 features for a single accelerometer and 44 features for two accelerometers; in both cases one feature represented the recognised activity.

The models for the estimation of EE were trained using the Support Vector Regression (SVR) algorithm as implemented in the Weka machine-learning suite [22]. This algorithm outperformed other algorithms (Linear regression, REPTree, Multilayer perceptron) in our previous research on EE estimation [23]. All sensor evaluations were performed using the leave-one-person-out cross-validation.

The results of the evaluation are presented in Table III. The single-sensor evaluation shows that the accelerometer (ACC) yields the lowest error of estimation, thus becoming the first component of the sensor configuration for the method. Once the initial configuration was defined, the evaluation of different sensor sets was performed to find the best configuration for the estimation of EE when using one or two accelerometers. Since all non-accelerometer sensors could be integrated with the chest accelerometer in a single enclosure we denote the sensor combinations that could be worn on the chest as one enclosure, and those that require the thigh accelerometer as two enclosures.

By observing the results in Table III we can see that evaluation of additional sensors preferred the heart rate (HR) for inclusion, which is in accordance with the single-sensor results. The next best combination was with the near-body temperature, which performed poorly on its own. To clarify the reason for this, we analysed the correlations between the sensors. All the remaining sensors were well-correlated with the already included heart rate, so including them provided little additional information. The exception was near-body temperature, so even though it did not perform well alone, it provided some additional information to the model. From this experiment we concluded that in the case of a single model for the estimation of EE, we get the lowest mean absolute error (MAE) if we utilise the heart rate and near-body temperature in addition to one or two accelerometers.

By observing the results in Table III we can see that evaluation of additional sensors preferred the heart rate (HR) for inclusion, which is in accordance with the single-sensor results. The next best combination was with the near-body temperature, which performed poorly on its own. To clarify the reason for this, we analysed the correlations between the sensors. All the remaining sensors were well-correlated with the already included heart rate, so including them provided little additional information. The exception was near-body temperature, so even though it did not perform well alone, it provided some additional information to the model. From this experiment we concluded that in the case of a single model for the estimation of EE, we get the lowest mean absolute error (MAE) if we utilise the heart rate and near-body temperature in addition to one or two accelerometers.

TABLE III
RESULTS OF THE METHODS FOR ENERGY EXPENDITURE ESTIMATION EXPRESSED IN MAE FOR ONE AND TWO ACCELEROMETER ENCLOSURES USING SINGLE REGRESSION MODEL.

Configuration	Enclosure	
	One (MAE)	Two (MAE)
Single sensor evaluation		
ACC	**0.73**	**0.58**
HR	1.11	/
NBT	2.04	/
ST	1.88	/
GSR	1.89	/
deltaT	1.79	/
Inclusion of one additional sensor (ACC +)		
HR	**0.69**	**0.56**
NBT	0.71	0.58
ST	0.72	0.58
GSR	0.72	0.58
deltaT	0.70	0.57
Inclusion of next additional sensor (ACC+ HR +)		
NBT	**0.66**	**0.54**
ST	0.67	0.56
GSR	0.70	0.57
deltaT	0.69	0.56
Inclusion of next additional sensor (ACC+ HR + NBT +)		
ST	0.66	0.55
GSR	0.67	0.55
deltaT	0.67	0.55

B. Choice of regression models

The analysis of the error of the regression models from the previous subsection suggested that multiple regression models utilising different sensors sets would improve the performance. For example, additional sensors (ACC + HR + NBT) improved the performance of the estimation of EE for moderate and vigorous activities compared to accelerometer only (ACC), and worsened it for light activities. This is shown as box plots in Fig. 2 in terms of the mean percentage error (MAPE), which is the MAE divided by the true EE. The results for these two models, when used for all the activities, are shown in Table V as “One model”.

Table IV presents the intensity of physical activity (light, moderate, vigorous) with the range in MET, and the activities that are recognised for each intensity depending on the number of enclosures. In the case of two enclosures ten activities are

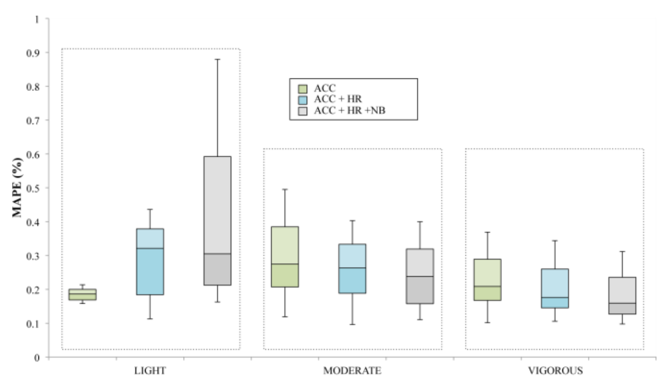

Fig. 2. Comparison of different single models for estimation of EE.

recognised, while in the case of one enclosure only seven activities are recognised. The reason for this is that standing and siting cannot be reliably distinguished with one accelerometer and are thus merged into the upright-position activity; kneeling, leaning and all fours are merged into the dynamic activity.

Dividing the activities by intensity and using two models for the estimation, the accelerometer-only model (ACC) for light activities and the model with additional sensors (ACC + HR + NBT) for moderate and vigorous activities improved the results in both MAE and MAPE. This is shown in Table V as "Two models".

Further error analysis suggested that some activities benefited from their own model, while the EE for others was more accurate when a model for multiple activities was used. However, since overspecialized models incur the risk of poor performance on unexpected activities, we find joining similar activities preferable. This also increases the amount of training data available for each model. For example, running and walking are similar and fairly distinctive activities with a large range of possible EE values. This makes them good candidates for a single model, since such a model can interpolate and extrapolate well from the intensities of the walking/running in the training data to the whole range.

Four models were trained for one sensor enclosure and four models for two sensor enclosures. The decision on which model to use is done in two stages, both relying on the recognised activity, as shown in Fig. 3: first, according to the current intensity of activity, and second, according to whether the recognised activity requires a specific model trained only on the data of that activity. Light activities use models that utilise only the acceleration data: M_{1L} in for one enclosure and M_{2L} for two enclosures. Moderate and vigorous activities use models that also utilise the data from additional sensors: M_{1MV} for one enclosure and M_{2MV} for two enclosures. The sensor combinations for the activity-specific models are the same as for the moderate and vigorous activities. They are used in the following cases:

- The recognised activity is dynamic in the case of one enclosure: M_{1D}.
- The recognised activity is standing in the case of two enclosures: M_{2S}.
- The recognised activity is walking or running in both enclosure configurations: M_{1WR} and M_{2WR}.

The results for our final proposed method are visually presented in Fig. 4, and in terms of MAE and MAPE in Table V as "Our proposed method".

Finally, we compared our results against the work by Altini et al. [7] and Gjoreski et al. [8] and the BodyMedia Fit EE estimation armband. While many other methods exist, these three were chosen because the method by Altini et al. was shown to perform best out of three common approaches to EE estimation, the method by Gjoreski et al. probably uses the most advanced algorithm (albeit not a lot of domain knowledge), and the BodyMedia Fit armband is the most advanced dedicated consumer device for EE estimation.

We re-implemented the method by Altini et al. [7].

TABLE IV
INTENSITY OF PHYSICAL ACTIVITY ACCORDING TO THE RESULT OF ACTIVITY RECOGNITION.

Intensity of physical activity	Recognised activity: 1 Enclosure	Recognised activity: 2 Enclosures
Light $0.9 \geq MET < 3$	Lying	Lying
	Upright position	Sitting, Standing
	Dynamic	Kneeling, Leaning, All fours
Moderate $3 \geq MET \leq 6$	Transition, Walking	Transition, Walking
Vigorous $MET > 6$	Running, Cycling	Running, Cycling

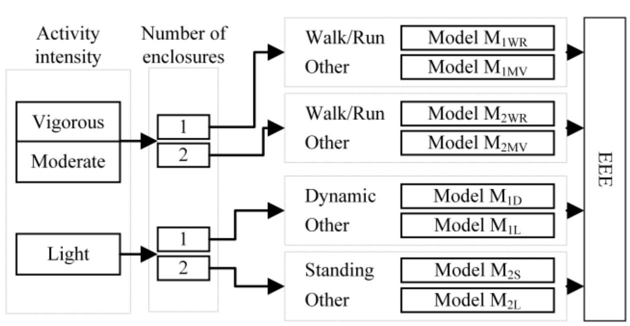

Fig. 3. Workflow of our proposed method.

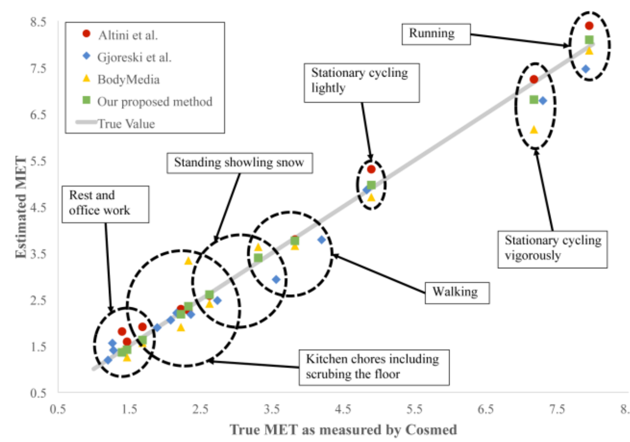

Fig. 4. Comparison of estimation for all competing methods.

TABLE V
COMPARISON OF PROPOSED METHOD FOR ESTIMATION OF ENERGY EXPENDITURE TO THE COMMERCIAL SYSTEM BODYMEDIA.

Method	One Enclosure MAE (MET)	One Enclosure MAPE (%)	One Enclosure RMSE (kcal/min)	Two Enclosures MAE (MET)	Two Enclosures MAPE (%)	Two Enclosures RMSE (kcal/min)
One model for all activities						
ACC	0.73	27.0	1.37	0.58	26.7	1.14
ACC + HR + NB	0.66	28.3	1.26	0.54	26.3	0.96
Multiple models						
Two models	0.64	25.3	1.01	0.54	23.5	0.86
Our proposed method	**0.62**	**24.6**	**0.91**	**0.52**	**23.2**	0.81
Compared to						
Altini et al. [7]	**0.73**	**26.3**	**0.96**	**0.70**	**25.7**	**0.88**
Gjoreski et al. [8]	-	-	-	**0.61**	**27.5**	-
BodyMedia Fit	**0.87**	**32.8**	**1.35**	-	-	-

Following the authors' philosophy, we divided the activities by scenario into two categories: sedentary (lying, sitting, standing) and active (walking, running, cycling and HWBM). For each sedentary activity, we trained a regression model that computed per-user MET lookup values, and for each active activity we trained an activity-specific model using acceleration and heart rate data. The results show that our proposed method outperforms the one by Altini et al. by 0.11 MET or 2.7 percentage points in the case of one enclosure, and by 0.18 MET or 2.5 percentage points in the case of two enclosures. Additionally, we compared our proposed method to the one by Altini et al. in terms of RMSE using kcal/min instead of MET, since this was the measure used by Altini et al. Our method outperformed the one by Altini et al. by 0.05 kcal/min in the case of one enclosure and by 0.07 kcal/min in the case of two enclosures.

The comparison of our proposed method against the one by Gjoreski et al. [8] can be done without re-implementation because the authors used the same dataset (*E* dataset) as we did for the experiments. Since their research used two accelerometer enclosures, we can only compare the results where we used two enclosures. We can observe that our method outperforms the one by Gjoreski et al. by 0.09 MET or 4.3 percentage points. We can observe that our method outperforms the BodyMedia Fit by 0.25 MET or 8.2 percentage points or 0.44 kcal/min.

The results of our and the competing methods are also shown in Fig. 4. We can observe that method by Altini et al. slightly overestimates the sedentary activities such as lying and standing. Other activities benefitted from activity-specific models, except for running and stationary cycling lightly, which were overestimated. The context-based method by Gjoreski et al. tends to underestimate exercise activities and overestimate the sitting activity with additional movement (office work). The reasons might be the poor exploitation of the accelerometers (only one feature in addition to the recognised activity) and indiscriminate use of all the sensors (e.g., higher heart rate during sitting does not always indicate higher EE). The BodyMedia Fit armband is intended for physically active people engaging in sports, which may explain the underestimation of light activities – when similar activities are a part of sports, they are performed more intensely. Interestingly, the BodyMedia Fit armband also underestimated running and stationary cycling. Our method is quite accurate for all activities. Its error was largest when cycling vigorously, where the approach by Altini et al. performed best.

We also calculated the statistical deviation and two-tailed t-test for statistical significance for the methods for which we obtained the estimations. The average and the standard deviations are presented in Table VI.

C. Validation of the proposed method

For the validation of the proposed method, we used the *V* dataset, containing the recordings of five different people performing slightly different activities than the *E* dataset. We used the models for EE estimation build on the *E* dataset and applied them on the *V* dataset. The results can be observed in Table VII. Compared to evaluation results, we can see that the MAE is slightly higher and MAPE is lower in both enclosure cases. To understand this discrepancy, we compared the estimation errors on both datasets per activity. The main difference was a larger MAE in the case of running on the *E* dataset, while the other errors on the *E* dataset were smaller or comparable. Since the EE in the case or running is large, a modest increase in MAE translated to a small increase in MAPE, which was compensated by the other errors being smaller (Fig. 5.). We can observe that range of error box of estimation of EE in walking and sitting activity expands in validation dataset, due to additional scenarios compared to evaluation dataset. However, the error does not exceed the boundaries set by evaluation dataset. The comparable results on the *E* and *V* datasets show that our method did not overfit the training data.

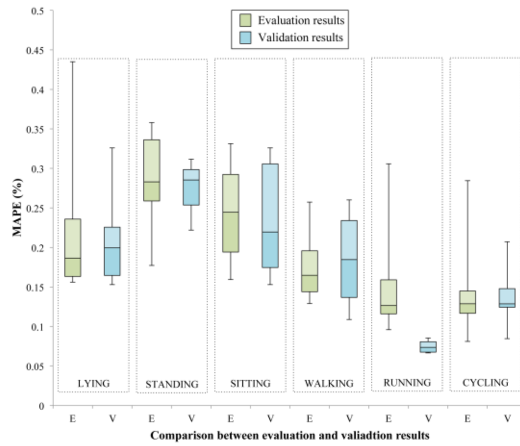

Fig. 5. Comparison of estimation error (MAPE) per activity (E – evaluation dataset, V – validation dataset). Presented results are for two enclosures.

TABLE VII
COMPARISON OF PROPOSED METHOD ON BOTH DATASETS

Models	One Enclosure		Two Enclosures	
	MAE (MET)	MAPE (%)	MAE (MET)	MAPE (%)
Evaluation results	0.62	24.6	0.52	23.2
Validation results	**0.69**	**22.5**	**0.56**	**19.4**

VI. CONCLUSION AND DISCUSSION

This paper presents an approach to designing a method for EE estimation using multiple sensors and regression models. The approach first evaluates different sensors and their combinations. It then considers which activities should have their own dedicated regression models, and which should use common models. Each of the models may also use its own

TABLE VI
STANDARD DEVIATIONS AND STATISTICAL SIGNIFICANCE

	Altini et al.	BodyMedia Fit	Our method
Average (MAE)	0.70	0.87	0.52
STD (*p-value*)	±0.19 ($p<0.05$)	±0.33 ($p<0.02$)	±0.15

sensor set. Our method for EE estimation designed in this way outperformed state-of-the art competing methods by up to 8.2 percentage points in terms of the MAPE.

In the course of designing the EE-estimation method, we confirmed that the recognised activity indeed plays an important role in EE estimation: in addition to being an important feature, it also gives a context of physical intensity (light, moderate, vigorous) on which the selection of sensors is based. Experiments showed that for light activities, the EE-estimation models should utilise only acceleration and no additional sensor data. This is probably because light activities are usually accompanied by normal heart rate and normal temperature, so these two parameters do not contribute any valuable information to the model, but they may contribute misleading information (increased heart rate because of a psychological trigger). Moderate and vigorous activities benefit from additional information such as the heart rate and near-body temperature, because these two parameters change with the intensity of activity, which is important if the activity has a large range of possible EE values. This is particularly valuable for cycling, where the movement does not necessarily reflect the effort, since it depends on the setting (gear) of the bicycle. The sensors for the skin temperature and galvanic skin response do not provide any additional information that can improve the model, probably because of their high correlation with the heart rate.

While the division of activity by intensity greatly contributes to accurate estimation, sometimes this is not enough. When the EE of an activity has a large range of values, for example walking and running combined, or standing comprising being still, cooking and cleaning, the amount of training data for each intensity is very small. If such an activity is joined with others when building a model, its data for each intensity is overpowered by all the other data, making the model inaccurate for that activity. Such activities require specialised models.

Our method uses wearable sensors that are attached to the user's body at a fixed location. This makes it suitable for dedicated devices for EE estimation. In the future, however, we intend to extend it to smartphone sensors (particularly accelerometer). This is challenging because a phone can be carried in various locations and orientations, but preliminary results on activity recognition and EE estimation using a smartphone independently of location and orientation yielded encouraging results. A phone can be used on its own or in combination with a consumer sensing device such as the Zephyr BioHarness chest strap or the FitBit device [24]. This will increase the applicability of our research, since both smartphones and consumer sensing devices are increasingly popular. In addition to transferring the method to more accessible sensors, we plan to research how to automatically adapt it to specific users [25].

Acknowledgment

This work has been partially supported by the EU project COMMODITY12 (www.commodity12.eu).

References

[1] S. B. Cooper, S. Bandelow, M. L. Nute, J. G. Morris, M. E. Nevill, "The effects of a mid-morning bout of exercise on adolescents' cognitive function," *Mental Health and Physical* Activity, vol. 5 no. 2, pp. 183–190, 2012.
[2] M. Hammer, E. Stamatakis, "Objectively assessed physical activity, fitness and subjective wellbeing," *Mental Health and Physical Activity*, vol. 3, no. 2, pp. 67–71, 2010.
[3] H. W. Kohl, C. L. Craig, E. V. Lambert, S. Inoue, J. R. Alkandari, G. Leetongin, S. Kahlmeier, "The pandemic of physical inactivity: global action for public health," *The Lancet*, vol. 380, no. 9838, pp. 294–305, 2012.
[4] World Heath Organisation, http://www.who.int/
[5] R. Brychta, E. Wohlers, J. Moon, K. Chen, "Energy Expenditure: Measurement of Human Metabolism," *Engineering in Medicine and Biology Magazine, IEEE* , vol.29, no.1, pp.42–47, 2010.
[6] BodyMedia Fit, http://www.bodymedia.com/armband.html
[7] M. Altini, J. Penders, R. Vullers, O. Amft, "Estimating Energy Expenditure Using Body-Worn Accelerometers: a Comparison of Methods, Sensors Number and Positioning," IEEE *Journal of Biomedical and Health Informatics*. Preprint, 2014.
[8] H. Gjoreski, B. Kaluža, M. Gams, R. Milić, and M. Luštrek. "Ensembles of multiple sensors for human energy expenditure estimation," In *Proc. of UbiComp* '13, 2013, pp. 359–362.
[9] S. E. Crouter, J. R. Churilla, D. R. Bassett Jr, "Estimating energy expenditure using accelerometers," *European Journal of Applied Physiology*, vol. 98, no. 6, pp. 601-612, 2006.
[10] S. E. Crouter and D. R. Bassett Jr, "A refined 2-regression model for the actigraph accelerometer," *Medicine & Science in Sports & Exercise*, vol. 42, no.5, pp. 1029—1037, 2010.
[11] D. Kim, H. C. Kim, "Estimation of activity energy expenditure based on activity classification using multi-site triaxial accelerometry," *Electronics Letters*, vol. 44, no. 4, pp. 266-267, 2008.
[12] C. W. Lin, Y. T.C. Yang, J. S. Wang, Y. C. Yang, "A Wearable Sensor Module With a Neural-Network-Based Activity Classification Algorithm for Daily Energy Expenditure Estimation," *Information Technology in Biomedicine, IEEE Transactions on*, vol.16, no.5, pp.991-998, 2012.
[13] M. Altini, J. Penders, and O. Amft, "Energy expenditure estimation using wearable sensors: a new methodology for activity-specific models", In *Proc. of the conference on Wireless Health*, 2012, pp.1–8.
[14] N. Vyas, J. Farringdon, D. Andre, J. I. Stivoric, "Machine Learning and Sensor Fusion for Estimating Continuous Energy Expenditure," *AI Magazine*, vol. 33, no. 2, pp. 55-66, 2012.
[15] J. M. Lee, Y. Kim, G. J. Welk, "Validity of Consumer-Based Physical Activity Monitors," *Med Sci Sports Exerc*, vol. 46, no. 9, pp. 1840-1848, 2014.
[16] Shimmer, http://www.shimmersensing.com/
[17] Zephyr BioHarness, http://www.zephyranywhere.com/
[18] Cosmed K4 b_2, http://www.cosmed.com/
[19] S. Kozina, H. Gjoreski, M. Gams, and M. Luštrek, "Three-layer activity recognition combining domain knowledge and meta-classification," *Journal of Medical and Biological Engineering*, vol. 33, no. 4, pp. 406–414, 2013.
[20] L. Breiman, "Random Forests," *Mach. Learn.* vol. 45, no. 1, pp 5–32, 2001.
[21] B. Cvetković, B. Kaluža, R. Milić, M. Luštrek, "Towards Human Energy Expenditure Estimation Using Smart Phone Inertial Sensors," in *Proc. of AmI* '13, 2013, pp. 94–108.
[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.
[23] E. Mazomenos, J. M. Rodríguez, C. Cavero Barca, G. Tartarisco, G. Pioggia, B. Cvetković, S. Kozina, H. Gjoreski, M. Lustrek, H. Solar, D. Marincic, J. Lampe, S. Bonfiglio, K. Maharatna, "Case Studies," *Systems Design for Remote Healthcare*, New York, Springer, 2014, pp. 277-332.
[24] Fitbit Flex, http://www.fitbit.com
[25] Cvetković, B. Kaluža, M. Luštrek, M. Gams, "Adapting Activity Recognition to a Person with Multi-Classifier Adaptive Training," *Journal of Ambient Intelligence and Smart Environments*. Accepted for publication, 2015.