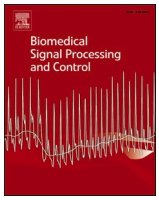
Biomedical Signal Processing and Control 87 (2024) 105381

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Dynamic personalized human body energy expenditure: Prediction using time series forecasting LSTM models

Victoria M. Perez Cortes, Arnab Chatterjee, Dolaana Khovalyg *

Laboratory of Integrated Comfort Engineering (ICE), École Polytechnique fédérale de Lausanne (EPFL), Fribourg, CH 1700, Switzerland

ARTICLE INFO

Keywords:
Human energy expenditure
Dynamic metabolism
Personalized model
Wearable sensing
Machine learning
Time series
Recurrent neural networks

ABSTRACT

Dynamic human energy expenditure (EE) consists of energy costs for resting metabolism, food digestion, physical activity, and thermoregulation. Currently, multiple models predict EE mainly concerning physical activity, thus, discarding other factors contributing to the dynamic variation of EE. This paper aimed to demonstrate that (i) a dynamic human body EE prediction requires the time series approach, (ii) personalization of input features and models can outperform the generalized approach. To achieve these objectives, data were collected from 3 sets of experiments with 6 test subjects wearing multiple sensors. The analysis of features' importance showed that the selection of features varies for activity-dominated and non-activity-dominated cases and also varies between individuals. Long-Short Term Memory (LSTM) networks were used to develop personalized models such as a simple LSTM, a convolutional LSTM (CNN-LSTM), and also an ensemble model combining CNN-LSTM with a Gradient Boosting algorithm (LSTM-LGBM). A personalized autoregressive linear model and a generalized approach of the LSTM-LGBM method were also developed to have a base of comparison. The results show that the personalized models provide good prediction accuracy, with the mean absolute percentage error (MAPE) mostly lying in the range of 5–15 %. The CNN-LSTM outperforms a simple LSTM model by 3–5 % in MAPE values, and the ensemble model outperforms the by 5–8 % the simple LSTM. The personalized modeling approach with LSTM has shown the potential to improve the prediction accuracy of dynamic EE and capture the non-activity-related effects such as thermoregulation and postprandial thermogenesis.

1. Introduction

Daily energy expenditure (EE) of the human body comprises four main parts: resting metabolic rate (RMR), the thermic effect of food (TEF), activity-induced thermogenesis (AT), and thermoregulation [1]. The RMR is the energy required to keep the human body functioning at rest, the TEF is the energy cost of chewing, swallowing, digesting, absorbing, and storing nutrients from food, and the AT includes the energy cost of physical exercise and non-exercise activity thermogenesis which ranges from simple standing to fidgeting and moving about [2]. For regular daily activities of people, non-exercise-induced thermogenesis is more profound [3,4]. While RMR in humans is mainly determined by body size, body composition, and genetic traits, the time scale is essential in TEF, similarly to AT, as postprandial EE is a transient phenomenon reaching its peak sometime (~15–120 min) after the meal intake depending on the type of food and individual's absorption capacity [5]. Thus, TEF and AT are dynamic components of EE, and they are highly individual as dependent on a person's habits and behavior (e.g., diet and physical activity choice). Finally, thermoregulation adjusts metabolic heat production in environments beyond neutrality due to insulative and metabolic adaptations [5,6], the main environmental determinant of EE is ambient temperature. Generally, dynamic human EE considering all contributing effects can be determined by measuring human body heat generation either directly using *direct calorimetry* [7] or by means of *indirect calorimetry* (measuring oxygen inhaled and carbon dioxide exhaled) [8]. However, these methods are mainly suitable for in-lab measurements as they require specialized high-cost bulky equipment. The advancement of electronics (hardware miniaturization and memory improvements) led to low-cost, lightweight, and power-efficient wearable sensors such as heart rate (HR) monitors, accelerometers, and gyroscopes, enabling easier estimation of EE using secondary biomarkers [9,10].

The recent systematic review and meta-analysis by O'Driscoll et al. [11] identified 40 different devices (1 forearm-worn, 6 worn on the upper arm, and 33 wrist-worn) predicting activity-related human EE using mainly accelerometry and HR measurements. A summary of

* Corresponding author.
E-mail address: dolaana.khovalyg@epfl.ch (D. Khovalyg).

https://doi.org/10.1016/j.bspc.2023.105381
Received 14 March 2023; Received in revised form 18 July 2023; Accepted 17 August 2023
Available online 7 September 2023

Table 1
Review of selected human energy expenditure (EE) prediction models and their performance.

Authors	Ref.	Study subjects characteristics				Activity types	Activity intensity	Personal data type collection*	Wearable sensors/ devices deployed	Algorithm**	Min. MAPE [%]	Min. RMSE	Main findings
		Number, sex	Age (y.o.)	Mean weight (kg)	BMI (kg/m^2)								
Hendelman et al. (2000)	[12]	25 subjects	30–50	n/a	n/a	Walking, golf playing, household tasks	Low-medium	ACC	CSA, Tritrac monitor	Regression equations	30	n/a	The relationship of accelerometry to energy cost is highly dependent on the type of activity being performed
Welk et al. (2000)	[13]	52 (21 M, 31F)	29	n/a	23.3	Choreographed routines	Lifestyle activities	ACC	CSA, Tritrac, Biotainer monitor	Regression equations	38	n/a	Accelerometry-based activity monitors provide less accurate predictions of EE
Swartz el al. (2000)	[14]	70 (31 M, 39F)	41± 15	76.2± 18.2	26.0 ± 4.5	Housework, yardwork	Lifestyle activities	ACC	CSA accelerometer	Regression equations	34	n/a	The combination of hip and wrist accelerometers improves EE prediction.
Spierer et al. (2011)	[15]	27 (16 M, 11F)	26.4	67.6	23.9	sweeping, lifting weights, walking, jogging	Low-medium	ACC	Actical	Linear regression	17	n/a	All estimates from Actical devices tended to underestimate EE
Dannecker et al. (2013)	[16]	19 (10 M, 9F)	26.9± 6.6	75.1± 17.1	25.1 ± 4.6	Sit, stand, walk, cycle	Low-high	ACC	Pressure sensing insole, Actical	Ordinary least squares linear regression model	–	10.4 % for sitting	Insole and foot-mounted accelerometer can accurately estimate EE during typical low-level (free-living) tasks
Zhu et al. (2015)	[17]	30 (20 M, 10F)	27.8 ± 6.9	66.9 ± 15.1	n/a	Walking, standing	Medium	ACC, HR, personal characteristics	Accelerometer, HR from COSMED K4b2	CNN, ANN	n/a	1.12 kcal/ min (CNN)	CNN matches most closely to the ground truth for EE prediction, while ANN diverges widely
Cvetkovic´ et al. (2016)	[18]	10 (8 M, 2F)	24–33	n/a	20–28.9	Sedentary, standing, light work, walking, running, cycling	Low-high	ACC, HR, BR, GSR, ST	Shimmer2 BodyMedia Fit	SVR	22.5	n/a	Activity recognition improves EE estimation. For light activities, using accelerometry data is sufficient to have a good EE estimate
Montoye et al. (2017)	[19]	39 (19 M, 20F)	22.1± 4.3	72.4± 16.2	24.4 ± 3.6		Low-high	ACC	GENEActiv accelerometers, ActiGraph	Linear regression, linear mixed models, ANN	n/a	1.07 MET (ANN)	ANN models developed to predict EE from features extracted from raw acceleration data can significantly improve prediction accuracy for wrist-worn accelerometers
O'Driscoll et al. (2020)	[20]	59 (18 M, 41F	44± 14.1	75.7± 13.6	26.9 ± 4.7	Sitting, standing, walking, jogging	Medium-high	ACC, HR, ST, GSR, personal characteristics	Fitbit charge 2, Polar H7, SenseWear arm band, Actigraph	Random Forest	18.3	1 MET	The models with more accelerometer variables as input features led to the greatest predictive accuracy, indicating the importance of tri-axial accelerometry.
Sevil et al. (2020)	[21]	25 (12 M, 13F)	24.9± 3.2	72.4± 14.2	25.0 ± 5.0	Sedentary, walking, light work, cycling, training, running	Low- high	ST, HR, ACC, GSR, BVP	Empatica E4 wristband	k-NN, SVM, DT, Naïve Bayes, GPR, linear discrimination, NN, EL, LSTM	12.1	0.51 MET	LSTM performs better than the other algorithms. The use of multiple sensors and a sensor fusion results in better EE estimation than a single sensor approach.

* ACC – accelerometry, HR – heart rate, BR – breath rate, BVP -blood volume pulse, GSR – galvanic skin response, ST – skin temperature, ECG – electrocardiography.
** SVM: support vector machines, k-NN: k nearest neighbors, DT – decision trees, CNN/ANN: convolution/ artificial neural network, GPR – Gaussian process regression, EL – ensemble learning, LSTM: long-short term memory model, GRU – grated recurrent unit.

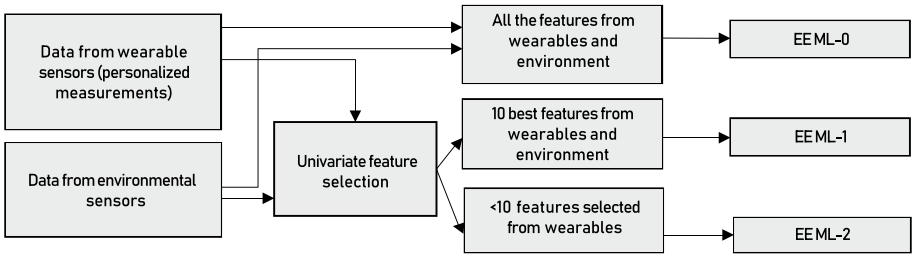


Fig. 1. Flowchart of the proposed data-driven energy expenditure models.

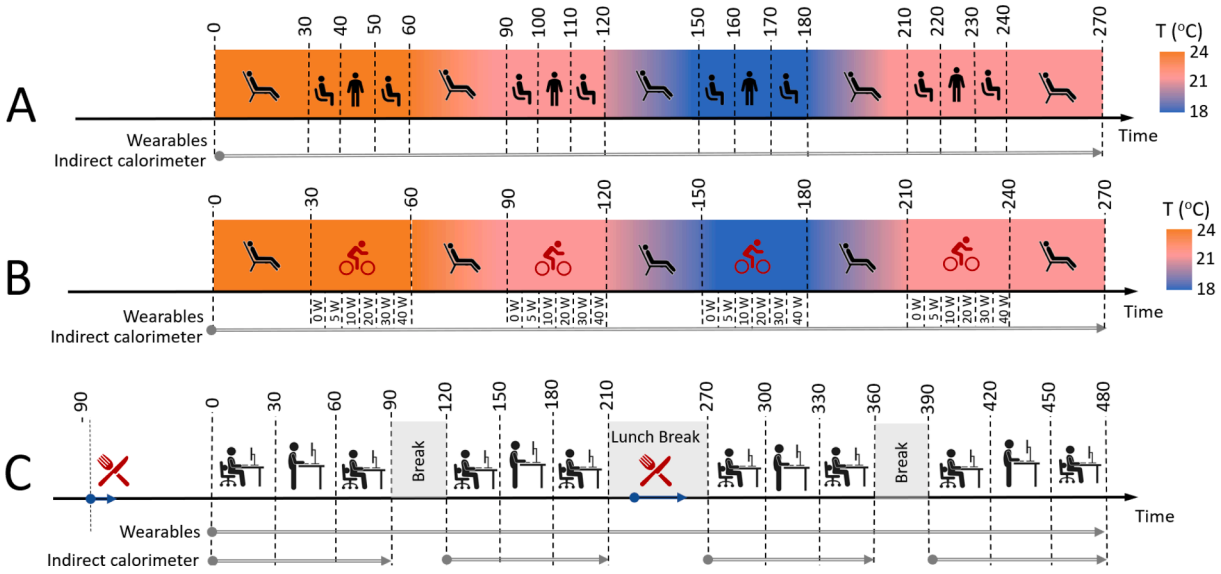


Fig. 2. Schematic of the design of three experimental protocols A, B, and C (time in minutes).

Table 2
Overview of the body accelerometry measurements (in counts) and their labeling.

Sensor type	Position	Label per component of accelerometry			Sampling frequency (Hz)
		x	y	z	
CALERAresearch (greenteg AG, CH)	lower chest	*ACCx_lchest*	*ACCy_lchest*	*ACCz_lchest*	1
gSKIN® BodyTemp Patch sensors (greenteg AG, CH)	upper chest	*ACCx_uchest*	*ACCy_uchest*	*ACCz_uchest*	1
	hand	*ACCx_hand*	*ACCy_hand*	*ACCz_hand*	1
Empatica E4 (Empatica, US)	wrist	*ACCx_wrist*	*ACCy_wrist*	*ACCz_wrist*	32
MOX 1 (Maastricht Instruments, NL)	thigh	*ACCx_thigh*	*ACCy_thigh*	*ACCz_thigh*	25

selected studies is provided in Table 1. The works [12–16] used only accelerometry data from various body parts fitted into the regression equations to estimate the EE. In addition, they focused on predictions per activity type, as the precision was improved by labeling posture and activity type. The use of Machine Learning (ML)-based models emerged in the last decade, and studies [17–20] deployed algorithms for classification and regression problems, such as support vector machine (SVM), decision tree (DT), random forest (RM), and neural networks (ANN, CNN). Recent work by Sevil et al. [21] also focused on a group of supervised ML algorithms to classify the physical activity type, but additionally the deep learning model Long-Short-Term Memory (LSTM) was also explored for minutely prediction of EE for a single participant. The LSTM model had the highest prediction among all ML algorithms tested, with a mean absolute percentage error (MAPE) of 12.1 %. Thus, amongst various ML algorithms, LSTM promises the highest accuracy in addition to the possibility to perform real-time continuous prediction rather than classification only.

Literature review shows that physical activity-related EE has been the main objective of the studies using a wide range of inertial sensors, sometimes with incorporated skin temperature sensors together with HR monitors. However, diet and environment-related factors had loose attention; the experiments were conducted at room temperatures (without detailing actual thermal exposure), the description of experimental studies lacked information on whether participants were fasted or fed, and the duration of experiments was rather short (less than 1 h) unable to capture the postprandial effect. To better pinpoint the effect of not only activity but also diet and thermoregulation on EE prediction, this work aims to demonstrate: (i) the prediction accuracy by considering parameters such as skin temperature and skin heat flux along with accelerometry and HR sensing; (ii) dynamic (real-time) EE prediction using time-series-based algorithms. As a time series-based algorithm, the LSTM is chosen as it could surpass other supervised learning methods. Advanced convolutional LSTM and ensemble LSTM were additionally explored, as only a few studies used this ensemble approach for an LSTM model [22,23]. As LSTM networks consider measurements from the past in their prediction [24,25], modeling EE on individual data sequences is more reasonable compared to the group approach mainly used in the prior works. As the variation in EE was shown to be subjective and

Table 3
Overview of human temperature and heat flux measurements and their labeling.

Sensor type	Position	Label per body side			Sampling frequency (Hz)
		Right	Left	Central	
Skin temperatures (°C)					
Empatica E4 (Empatica, US)	left wrist	–	*Tskin_wrist*	–	4
iButton® temperature loggers DS1922L (MAXIM Integrated, US)	forehead	–	–	*Tskin_forehead*	0.1
	neck	–	–	*Tskin_neck*	
	scapula	*Tskin_rscapula*	*Tskin_lscapula*	–	
	chest	–	*Tskin_luchest*	–	
	upper arm	*Tskin_ruarm*	*Tskin_luarm*	–	
	lower arm	*Tskin_rlarm*	*Tskin_llarm*	–	
	hand	–	*Tskin_lhand*	–	
	fingers	*Tskin_rfinger*	*Tskin_lfinger*	–	
	abdomen	*Tskin_rabdomen*		–	
	paravertebral	*Tskin_rpvbl*	*Tskin_lpvbl*	–	
	anterior thigh	*Tskin_rathigh*	*Tskin_lathigh*	–	
	posterior thigh	*Tskin_rpthigh*	*Tskin_lpthigh*	–	
	shin	*Tskin_rshin*	*Tskin_lshin*	–	
	calf	*Tskin_rcalf*	*Tskin_lcalf*	–	
	instep	*Tskin_rinstep*	*Tskin_linstep*	–	
Core body temperature (°C)					
CALERAresearch (greenteg AG, CH)	chest	–	–	*Tcore_chest*	1
Skin heat flux (W/m^2)					
gSKIN® BodyTemp Patch sensors (greenteg AG, CH)	*hand**	*Hfa_hand*	–	–	
		HFb_hand	–	–	1
	*chest**	–	*Hfa_uchest*	–	
		–	*HFb_uchest*	–	

*a sensor gSKIN® BodyTemp Patch has two channels, A and B; thus, there are two labels per location.

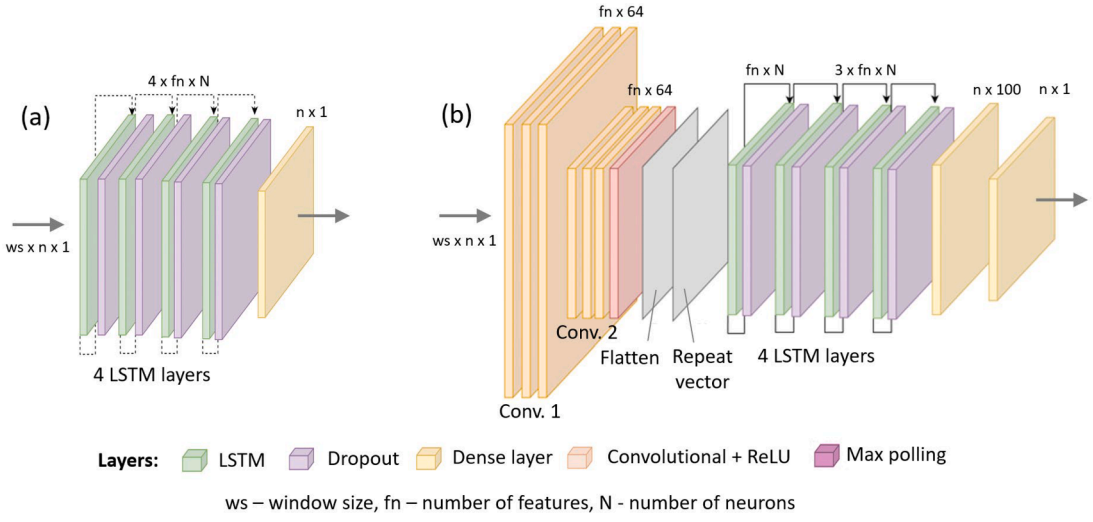


Fig. 3. Schematic of LSTM networks: (a) a simple LSTM network, (b) a convolutional LSTM network.

differing between individuals in the study by Khovalyg & Ravussin [26], the ultimate contribution of this work is to demonstrate the extent of a personalized approach to EE (i.e., developing the subject-specific modeling architecture and selection of features) might outperform the generalized approach.

2. Methodology

Several steps were involved in developing the data-driven prediction model of EE. The first is data collection, the second is features selection, and the final is modeling, including training and testing steps. The development of modeling was approached by 3 methods illustrated in Fig. 1. The most inclusive model, labeled **EE ML-0**, can be trained by including all the features from the wearables and environmental sensors. A model **EE ML-1** would be trained by including only the 10 best features from the wearables and environmental sensors using Univariate Feature Selection. Finally, the simplest and the most practical model **EE ML-2** would include only a minimum number of variables from only wearables, based on the top correlated features from the top 10 selected for EE-ML-1. For personalized prediction, feature selection was individualized, while for the generalized model, the dataset from all participants was combined. The following sections present in detail the methods used in each step.

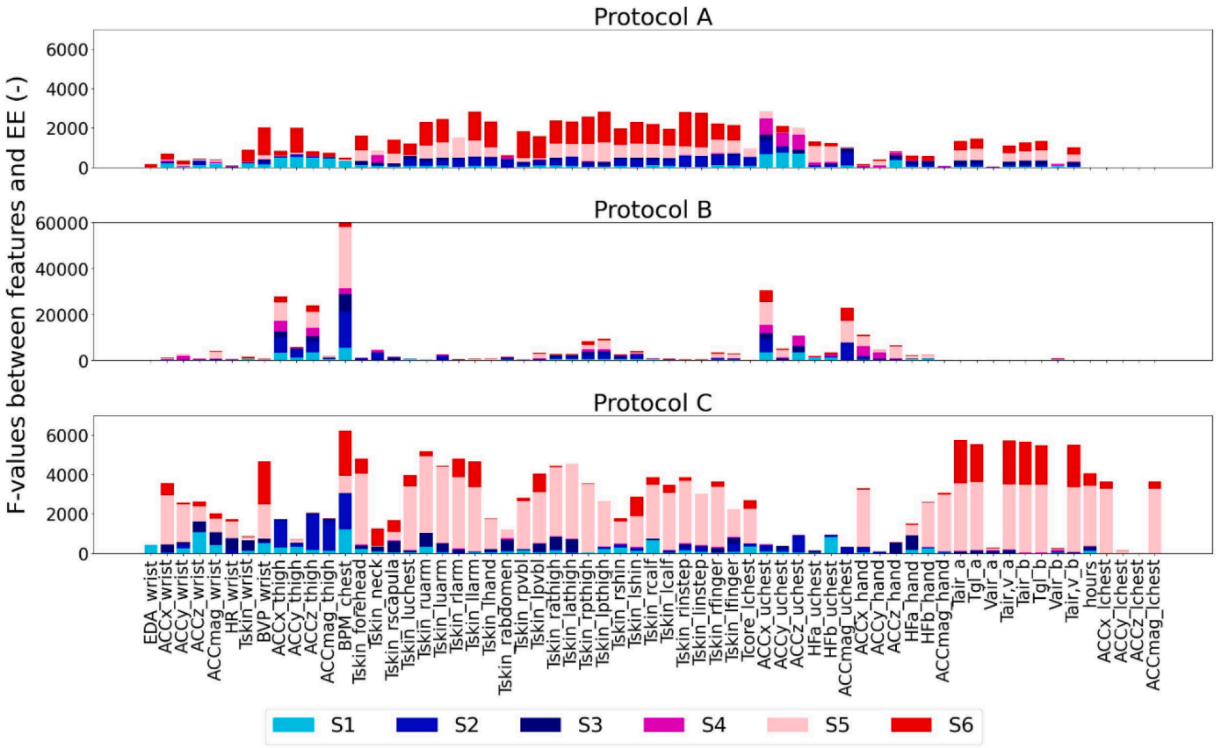


Fig. 4. F-values between features and EE across protocols and participants (stack plots).

2.1. Experimental design

To collect the required data, specially designed experiments were performed in 2020–2021 with three males (subjects S1-S3) and three females (subjects S4-S6). The average characteristics of participants were 30.2 ± 3.4 y.o., BMI 23.2 ± 3.4 kg/m^2; a detailed overview of participants and their body characteristics is provided in Khovalyg & Ravussin (2022) [26]. The study was approved by the Commission for the Ethics of Research on Human Beings (project 2020–02534). Three different protocols (A, B, and C) were conducted, considering varying activities, feeding status, and environmental temperature. Each protocol had 4 repetitive sessions in terms of physical activity. The overview of activities and environmental temperature for each protocol is provided in Fig. 2. The duration of Protocols A and B was 4.5 h (8:00–12:30), and they took place in the climatic chamber of the ICE lab at EPFL [27]. Protocol C was a full-day experiment (8:30–16:30) in the glazed facility CELLS (a shared building prototype at Smart Living Lab, Fribourg) [28]. The experiments were conducted with one participant at a time. The design details of each protocol are the following:

- The objective of Protocol A was to consider the variation in EE in test subjects during low-level isometric movements such as posture maintenance for 10 min (i.e., sitting and standing as most commonly encountered during everyday activities) at varying temperatures between 18 and 24 °C in each session.
- The design of Protocol B was similar to Protocol A, but the activity considered was graded low-power cycling on an ergometer, a standardized variation of walking about at a low to medium pace. Participants cycled at 60 rpm for 5 min at no-load (0 W) followed by 5, 10, 15, 20, 30, and 40 W of mechanical power output for 5 min each. Participants remained in a semi-reclined relaxed position when the temperature was changing from one set point to another, e.g., from 20 °C to 18 °C.
- The objective of Protocol C was to measure variation in the individual's EE and activity levels in test subjects during normal office work considering the meal effect. The subjects were required to consume a standardized breakfast (528 kcal) at 7:00, and a standardized lunch (884 kcal) at noon. Starting 8:30, there were four 1.5-hour sessions, and each session had alternating sitting and standing work. There were breaks at 10:00 and 14:30 when subjects could have some refreshments and use the restroom. The indoor temperature in Protocol C was not strictly maintained as in Protocols A and B, it varied in the range of 22–27 °C due to internal and solar heat gains, and the drifting pattern was not consistent between subjects. Also, reference EE measurements using an indirect calorimeter were stopped during the breaks. A detailed description of Protocol C is provided in [29].

2.2. Data collection

During the experiments, parameters related to different dimensions of the thermal environment, such as air temperature (*Tair*), globe temperature (*Tgl*), and airspeed (*Vair*), were measured at 1.1 height in the vicinity of the participant at a 1 s rate to characterize the environment surrounding him. The location "*a*" was at the back of the person, "*b*" was on the person's left. Apart from environmental sensing, participants were heavily sensed using multiple wearable sensors. The reference EE was measured using an indirect calorimetry Quark CPET (Cosmed, IT) with a silicon facemask system. The calorimeter measured the volume of inhaled O2 (*VO2* in *ml*/min) and exhaled CO2 (*VCO2* in *ml*/min) on a breath-to-breath basis. From measured *VO2* and *VCO2*, EE in *kcal*/min is calculated according to the Weir equation $EE = 3.9 \bullet VO2 + 1.11 \bullet VCO2$ [30]. Heart rate in *bpm*, labeled as *BPM_chest,* was measured using Actiheart 5 (Camntech, UK) sensor. A wristband Empatica E4 (Empatica, US) worn on a non-dominant hand, it measured multiple parameters such as blood volume pulse (*BVP_wrist* in *mV* at 64 Hz*),* electrodermal activity (*EDA_wrist* in μS at 4 Hz), heart rate (*HR_wrist* in *bpm),* skin temperature (*Tskin_wrist*), and body accelerometry. Accelerometry of a particular body part was also measured by other sensors such as CALERAresearch (greenteg AG, CH), gSKIN® BodyTemp Patch sensors (greenteg AG, CH), and MOX 1 (Maastricht Instruments, NL). All measured components of accelerometry at different body parts are listed

Table 4
Top ten features selected for models EE ML-1 and EE ML-2 based on F-values according to their rating (numbering according to the importance, marked with (*) – features for EE ML-2 out of 10 included in EE ML-1)

Model type		Personalized EE-ML1 and EE-ML2																		Generalized		
Protocol		A						B						C						A	B	C
Subject ID		S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6			
Environment	Tair_a																		3*			
	Tgl_a																	8	8			
	Vair_a																		2*			
	Tair_b																	10	4*			
	Tgl_b																	9	7*			
	Vair_b																		6*			
Skin Temperatures	Tcore_chest																			6		5*
	Tskin_forehead			5*											9			5*				
	Tskin_neck			8*	4				9		10									7		
	Tskin_rscapula									6					8	5*				5*		
	Tskin_luchest			6*																4*		
	Tsin_ruarm													9		2*		1*				
	Tskin_luarm					7	10									10		2*				
	Tskin_rlarm		10			1*									4			4*				
	Tskin_llarm		8			6	4*												9			
	Tskin_lhand		9				8*															
	Tskin_wrist						5*							5*					5*			
	Tskin_rfinger		6*			10									2							
	Tskin_lfinger		5*					9							3		6					
	Tskin_rabdomen														1	8				1*		9
	Tskin_rpvbl						7*													3*		
	Tskin_lpvbl																			2*		
	Tskin_rathigh					9												6*				
	Tskin_lathigh		7*													4*		3*				
	Tskin_rpthigh			10		3*	6*		8	7	9		6*					7		9	5*	10
	Tskin_lpthigh					2*	3	10	7*			8*	7*								6	
	Tskin_rshin									8				10							7	
	Tskin_lshin					8			10	10					6				10		8	
	Tskin_rcalf													4*								
	Tskin_lcalf																					
	Tskin_rinstep		3*				2*								5							
	Tskin_linstep		4*				1*								7		10					
Heat Flux	Hfa_hand			9												1*						
	HFb_hand			7*																		
	Hfa_uchest				6	4*		8					10									1*
	HFb_uchest					5*		7			8*		8*	3*								6
Accelerometry	ACCx_uchest	3*	1*	1*	1*			3*	4*	4*	4*	2*	3*		10		7				2*	
	ACCy_uchest	1*			3					9		9								8		7
	ACCz_uchest	2*	2*	3*	2*			2*		1*	1*						5*					
	ACCmag_uchest			2*					3*			3*	2*				8					4*
	ACCx_hand				10						2*	7*					9				9	3*
	ACCy_hand				8						6*											
	ACCz_hand	9		4*	5							6*				9						
	ACCmag_hand																					2*
	ACCx_wrist																				3*	
	ACCy_wrist				9									2*		7						
	ACCz_wrist											10		7							4*	
	ACCmag_wrist															3*						
	ACCx_thigh	7*						4*	2*	2*	3*	4*	5*				4*					
	ACCy_thigh	5*					9	6	6*					8								
	ACCz_thigh	6*						1*	5*	3*	5*	5*	4*				2*					
	ACCmag_thigh	8															3*				10	
Others	HR_chest																					
	HR_wrist												9			6						
	BVP_wrist													6								
	EDA_wrist	10																				
	BPM_chest	4*						5*	1*	5	7*	1*	1*	1*			1*		1*	10	1*	

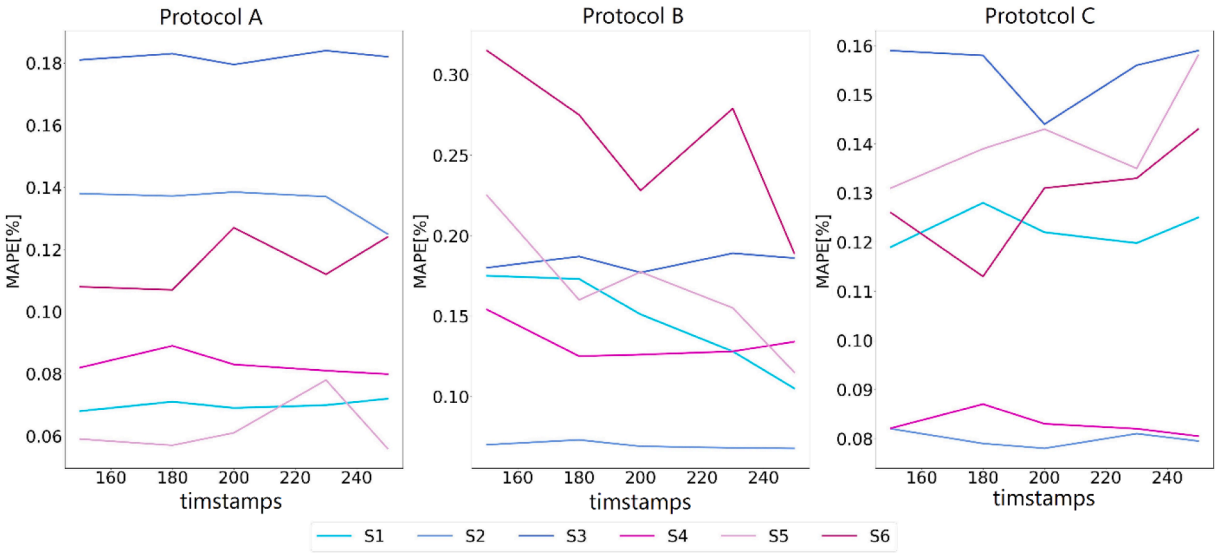


Fig. 5. MAPE variations for LSTM with changes in the number of previous timesteps for different subjects and protocols.

in Table 2. Skin temperature at 25 additional sites was measured using iButton® temperature loggers DS1922L (MAXIM Integrated, US); their labeling is provided in Table 3. Wearable sensors CALERAresearch also estimated core body temperature (*Tcore_chest*) using a patented algorithm when the sensor is placed on the side of the chest. Heat flux through the skin at selected locations was measured using gSKIN® BodyTemp Patch sensors, the labeling of measured locations is also listed in Table 3.

2.3. Univariate feature selection

More than 60 variables were measured from each human subject experiment; thus, only the features that have a reasonable correlation with the EE value should be selected to use as input variables in the prediction models EE ML-1 and EE ML-2 (Fig. 1) to reduce the computational cost of modeling. The ANOVA (Analysis of Variance) method, popularly known as the Univariate Feature Selection, was applied to select the top features using *scikit-learn* Python library. ANOVA works by selecting the features with the top scores based on univariate statistical tests where each feature is compared to the target variable to see whether there is any statistically significant relationship between them. The other option was to use PCA (Principal Component Analysis), but PCA does not estimate specific effects; it simply finds the mathematical definition of the "best" components (components that maximize variance). On the other hand, ANOVA estimates each component's specific factors, which was important in this study to find the feature set for the most practical EE prediction model.

2.4. Prediction model with LSTM

The prediction modeling of EE implemented was based on LSTM, unlike predictive regression modeling, adding the complexity of a sequence dependence among the input variables. The first step was data cleaning (replacing missing values, outliers, and noise removal using a rolling average) to transform the raw data into a machine-readable format. As different types of instruments were involved in the experiments, recording data at different frequencies, the data was synchronized to the same timesteps of 5 s. The dataset consisted of 3000–3500 data points (varied between subjects) for Protocols A and B, and 5950 in Protocol C. Once the preparatory steps were completed, the data was divided into training and testing parts. As we deal with time series, instead of using k-fold cross-validation, approximately ¾ of the data set was used for training and the remaining ¼ was used for testing. The exact splitting was based on the protocol's specifics to ensure that 3 out of 4 repetitive parts (sessions) of each protocol presented in Fig. 2 were considered in the training set, and the last part was used for the testing set. For the generalized model, in each protocol, all the subjects except one were used for training, and the arbitrary subject left out was used for the validation. Five different modeling approaches were tested in this work, four of them were personalized, and one generalized:

1. **Baseline (p)** – simple autoregressive linear model, personalized
2. **LSTM (p)** – LSTM model, personalized
3. **CNN-LSTM (p)** – convolutional LSTM (CNN-LSTM) model, personalized
4. **Ensemble (p)** – ensemble model LSTM-LGBM (CNN-LSTM model with the Light Gradient Boosting Machine), personalized
5. **Ensemble (G) –** ensemble model LSTM-LGBM, generalized

First, **a simple autoregressive linear model** serving as a baseline was built, then, **LSTM** and **convolutional LSTM (CNN-LSTM)** models were developed. A special 3D data structure for LSTM was created, with the third dimension being the data of the variables in the past *n* timesteps (Fig. 3). The hyperparameters were tuned using the grid-search method on the ¾ of the data set, similar to the training approach. In this method, possible values for each parameter were defined, and then the models were trained and evaluated for each combination of values in the grid to select the one with the smallest MAPE. The following parameters were tuned using the grid-search approach: (i) a number of neurons (50, 100, 150, and 200), (ii) a number of epochs (50, 100, and 150, up to 1500 epochs were preliminarily tested, then reduced to the ranges where the loss was converging), (iii) mini batch size (32, 64, 100 and 128), (iv) window size of the input (120–300 timesteps corresponding to 10–25 min), (v) learning rate of the optimizer (0.1, 0.01, and 0.001). The activation function for the LSTM layers (Tanh for LSTM layers and ReLU for convolutional layers), the loss function (MSE), the optimizer (Adam) and its learning rate (0.01) were selected for all the models and remained then the same along each subject in each session. However, the rest of the hyperparameters were optimized for each subject and session, which led to personalized models because we trained on one subject at a time. The details regarding the properties of the final LSTM and CNN-LSTM models are provided in Appendix A. Finally, **an ensemble model (LSTM-LGBM)** was developed by keeping the optimized CNN-LSTM model and fine-tuning hyperparameters for LGBM model, aiming to improve the predictive performance by combining the strengths of multiple individual models. The Light

Table 5
Performance summary (MAPE in %) of all EE prediction models for all subjects.

Prot.	Sbj.	ML model	Baseline (p)	LSTM (p)	CNN-LSTM (p)	Ensemble (p)	Ensemble (G)	Best model
A	**S1**	**EE ML-0**	8.6	10.6	7.7	7.5	17.8	Ensemble (p)
		EE ML-1	7.4	10.5	7.8	8.0	11.2	Baseline (p)
		EE ML-2	5.6	9.0	8.1	7.8	21.5	Baseline (p)
	S2	**EE ML-0**	13.4	18.0	14.0	14.0	17.8	Baseline (p)
		EE ML-1	32.9	16.0	12.5	12.0	14.3	Ensemble (p)
		EE ML-2	22.8	16.2	12.5	11.6	15.1	Ensemble (p)
	S3	**EE ML-0**	18.2	35.6	22.2	17.9	13.8	Ensemble (p)
		EE ML-1	23.4	15.8	16.2	15.1	16.2	Ensemble (p)
		EE ML-2	20.7	17.0	16.6	14.8	15.9	Ensemble (p)
	S4	**EE ML-0**	6.8	7.5	6.2	5.8	12.1	Ensemble (p)
		EE ML-1	16.0	8.2	6.6	5.1	8.8	Ensemble (p)
		EE ML-2	6.0	6.1	6.6	5.4	21.1	Ensemble (p)
	S5	**EE ML-0**	4.9	6.6	5.6	5.3	26.2	Ensemble (p)
		EE ML-1	5.4	5.5	7.2	5.6	31.7	Baseline (p)
		EE ML-2	5.0	5.9	7.0	5.6	25.4	Baseline (p)
	S6	**EE ML-0**	10.8	11.3	12.0	9.5	10.7	Baseline (p)
		EE ML-1	33.6	21.0	12.6	10.6	14.0	Baseline (p)
		EE ML-2	23.3	32.0	13.4	10.2	11.9	Ensemble (p)
B	**S1**	**EE ML-0**	31.5	16.0	9.3	19.4	14.0	CNN-LSTM (p)
		EE ML-1	16.6	14.9	7.9	15.8	12.9	CNN-LSTM (p)
		EE ML-2	16.6	12.3	7.1	14.7	11.7	CNN-LSTM (p)
	S2	**EE ML-0**	41.6	14.5	10.2	11.8	17.3	CNN-LSTM (p)
		EE ML-1	16.0	9.1	7.3	06.3	10.6	Ensemble (p)
		EE ML-2	12.5	8.1	8.4	09.2	19.2	LSTM (p)
	S3	**EE ML-0**	35.9	14.8	19.8	16.1	34.9	Ensemble (p)
		EE ML-1	23.3	12.3	18.5	16.8	24.4	LSTM (p)
		EE ML-2	22.1	13.4	18.8	16.1	24.1	LSTM (p)
	S4	**EE ML-0**	39.5	13.8	10.3	12.2	29.7	Ensemble (p)
		EE ML-1	30.4	16.6	13.6	7.9	12.5	Ensemble (p)
		EE ML-2	29.0	13.2	7.7	7.1	19.8	Ensemble (p)
	S5	**EE ML-0**	37.9	18.4	9.3	7.0	13.2	Ensemble (p)
		EE ML-1	20.9	9.5	7.9	7.5	11.9	Ensemble (p)
		EE ML-2	11.9	9.1	7.1	7.0	14.5	Ensemble (p)
	S6	**EE ML-0**	32.8	32.0	20.6	14.5	34.5	Ensemble (p)
		EE ML-1	20.9	23.2	17.3	12.2	47.7	Ensemble (p)
		EE ML-2	15.9	18.6	17.5	12.4	38.8	Ensemble (p)
C	**S1**	**EE ML-0**	7.2	8.0	9.2	7.9	10.2	Ensemble (p)
		EE ML-1	9.0	10.0	8.0	7.7	10.5	Ensemble (p)
		EE ML-2	10.5	8.6	9.6	7.4	12.8	Ensemble (p)
	S2	**EE ML-0**	7.5	13.8	13.6	8.2	10.1	Baseline (p)
		EE ML-1	8.1	8.3	7.5	6.1	11.4	Ensemble (p)
		EE ML-2	7.1	9.3	8.0	6.2	11.7	Ensemble (p)
	S3	**EE ML-0**	9.4	12.4	12.9	10.5	11.8	Baseline (p)
		EE ML-1	11.9	13.8	14.5	11.8	14.6	Ensemble (p)
		EE ML-2	9.8	11.2	14.2	10.2	12.2	Baseline (p)
	S4	**EE ML-0**	11.2	10.2	12.8	10.7	16.0	LSTM (p)
		EE ML-1	20.9	10.5	11.2	10.4	16.1	Ensemble (p)
		EE ML-2	22.3	12.9	12.1	12.7	18.4	CNN-LSTM (p)
	S5	**EE ML-0**	22.3	15.1	14.5	15.0	33.3	Ensemble (p)
		EE ML-1	15.3	18.5	15.1	13.1	16.7	Ensemble (p)
		EE ML-2	15.2	18.8	16.7	14.1	21.1	Ensemble (p)
	S6	**EE ML-0**	9.2	16.7	13.0	9.9	22.7	Baseline (p)
		EE ML-1	12.0	15.0	14.2	11.2	18.1	Ensemble (p)
		EE ML-2	9.1	14.2	14.2	11.3	31.5	Baseline (p)

Gradient Boosting Machine (LGBM) uses tree-based learning algorithms; thus, we had to tune the learning rate (0.05–0.31), maximum depth of the decision trees (1–16), minimum child weight (1–8), and colsample by a tree (0.3–0.8). The LSTM-LGBM model was tested for a personalized and generalized approach. The tuned hyperparameters for the personalized approach are provided in Appendix A. After the models were trained, MAPE in percentage was reported to evaluate the performance of models. It is important to point out that for the LSTM and CNN-LSTM models, 10 runs of model training were executed as they are stochastic by nature, and averaged predictions of the ten training sets are reported. The number of repetitions (runs) was chosen based on the robustness and stability of the results, and also based on the optimal time criteria. The codes of models are available on the GitHub repository https://zenodo.org/record/7584890 [31].

3. Results

The raw personalized data collected had some degree of variability depending on the type of protocol and the activities performed. Examples of the accelerometry and skin temperature data for selected cases are presented in Appendix B. As the main focus of the work is the presentation of models, the results highlight the features selection across protocols and participants and the performance of the EE prediction models.

3.1. Univariate feature selection across participants and protocols

The F-values between the measured features and the target variable EE for every subject are demonstrated as the stacked bar chart in Fig. 4. Table 4 lists all the top features selected for the EE ML-1 and EE ML-2

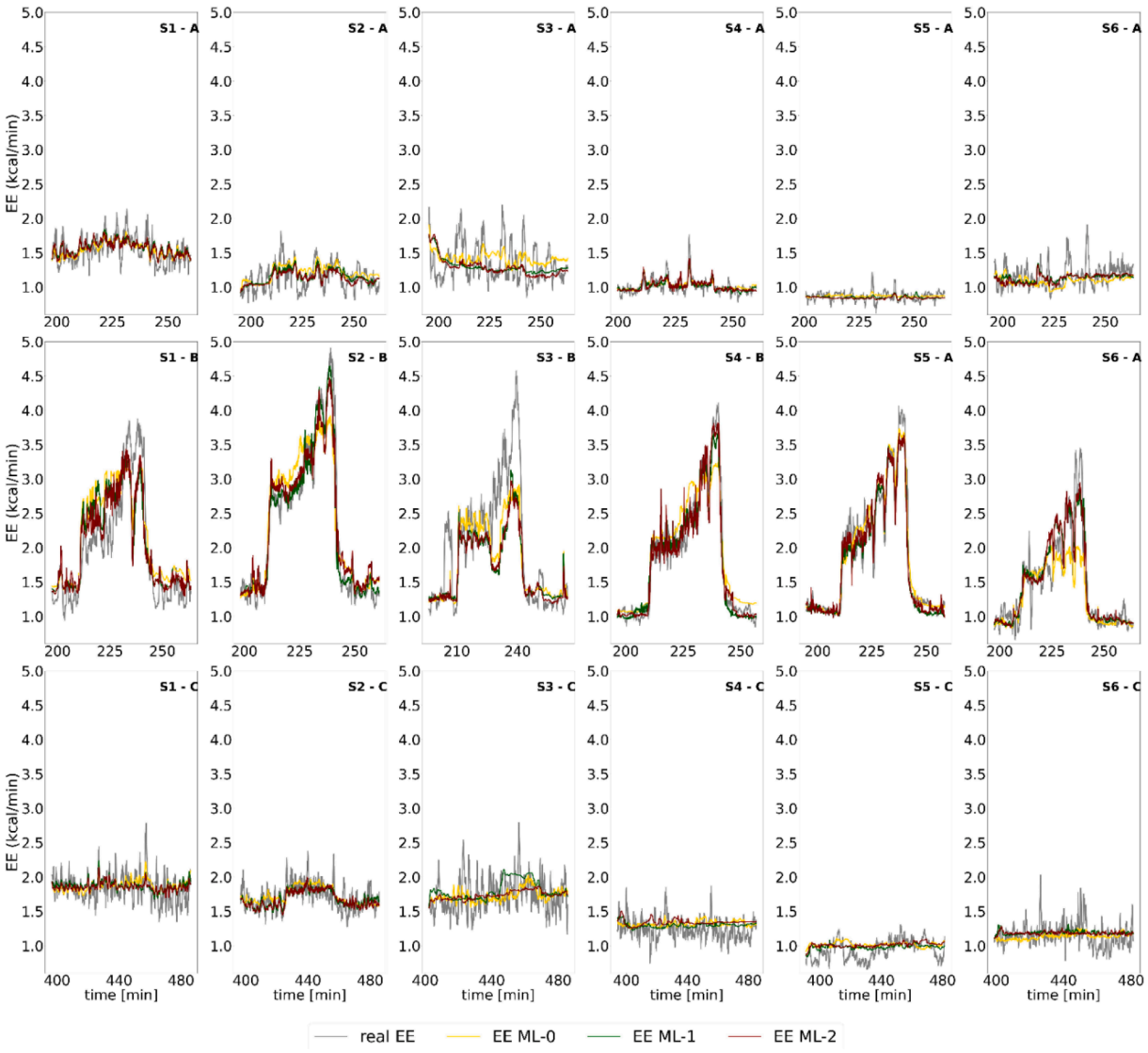


Fig. 6. Variation in actual EE and predicted EE using the ensemble model (LSTM-LGBM) for different subjects and protocols.

approaches (all features were considered in the EE ML-0). In Protocol A, with mainly upper chest movements, the accelerometry data from the upper chest and lower chest explain much of EE variance (particularly in participants S1-S4). Other than that, the skin temperatures from fingers, insteps, and thighs are also important, especially for females S5-S6. In the case of Protocol B with cycling activity, the heart rate measured at the chest *BPM_chest* has the highest or relatively high F-value for all subjects. In addition to that, due to the whole-body movement, accelerometry from the upper chest, thigh, and lower chest provides good correlations with EE. Finally, for Protocol C with typical standing-sitting office activities, the skin temperatures (mainly from arms, chests, calves, abdomen, and forehead) are among the most dominant ones. In a few participants, the heart rate from the chest *BMP_chest* and activity data from the thigh also exhibit high F-values. For medium to high-level physical activities (e.g., cycling), the heart rate and specific body part acceleration are important, and they overpower the skin temperatures effect; however, for normal day activities and the normal daily prediction of EE, the skin temperatures are important features for the EE prediction.

3.2. *Performance of ML-based EE prediction models*

As the LSTM algorithm is trained based on the variations in the data over a certain number of previous timesteps, timestamps between 120 and 300 (10–25 min) were assessed for each subject. Variation in MAPE for different numbers of previous timesteps shown in Fig. 5 indicates that the number of previous timestamps to reach the minimum MAPE is subject-dependent, strengthening the need for personalized EE prediction. The timesteps yielding the minimum MAPE were chosen for the final modeling of personalized models.

Performance in terms of MAPE of all models, personalized and generalized, is summarized in Table 5. In the case of Protocol A, all the models, except for the ones for S3, provide good prediction accuracy with MAPE of below 14 %. In the case of subject S3, the MAPE for only EE ML-0 is 18 %, while the EE ML-1 and EE ML-2 are between 14 and 15 %. For Protocol B, all the models, except those for subject S1, provide good prediction accuracies, with the MAPE value below 16 %. In the case of subject S1, the MAPE for only EE ML-0 is 19 %, while the EE ML-1 and EE ML-2 are between 14 and 15 %. Moving on to Protocol C, all personalized models provide excellent prediction accuracies, with the MAPE value below 12 %, except for the subject S5. Generally, personalized models perform better than generalized ones. Regarding the best

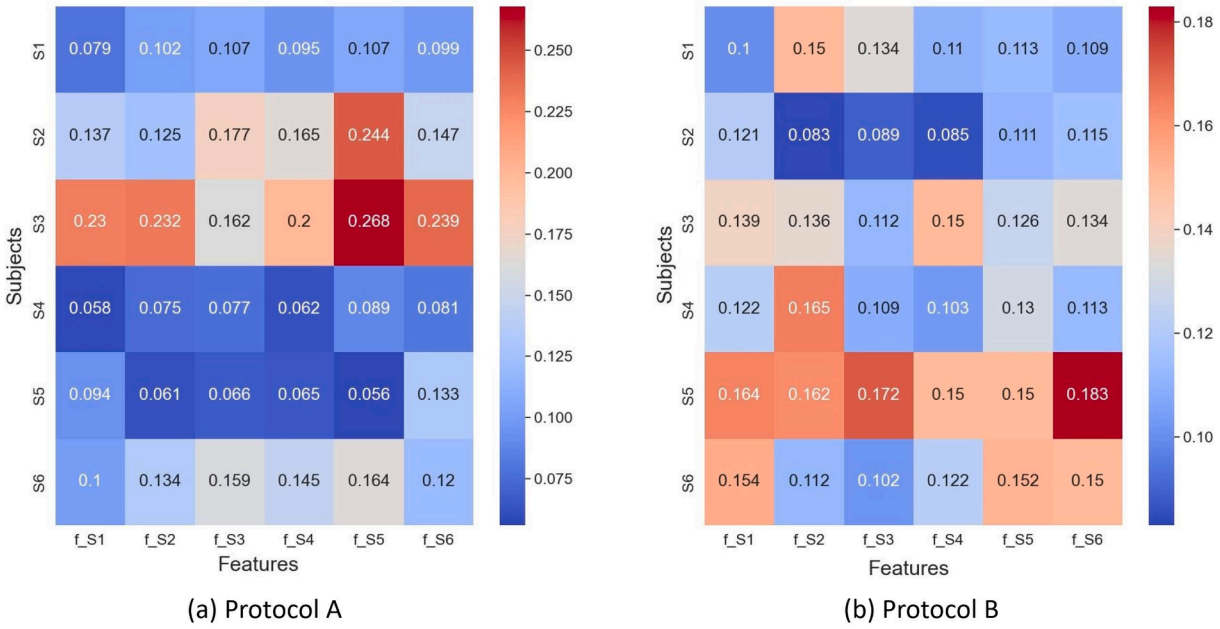


Fig. 7. Performance of the personalized ensemble model when personalized features are cross-used (MAPE values, to convert to % should be multiplied by 100): (a) Protocol A, (b) Protocol B.

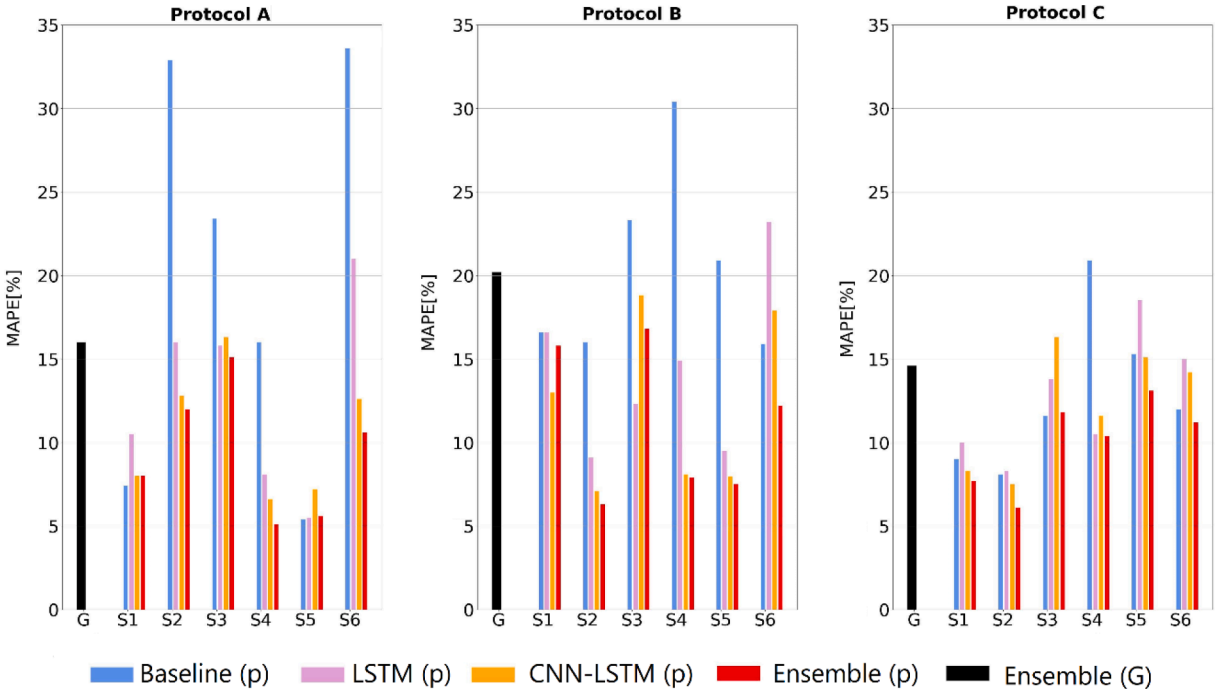


Fig. 8. Comparison of the MAPE for each model and subject for Protocols A, B, and C.

prediction accuracies considering all the protocols and subjects, the personalized ensemble model LSTM-LGBM generally performs better in 62 % of cases. It outperforms 10/18 cases in ML-0 and ML-2 approaches and 13/18 cases in ML-1 approach. The simpler baseline autoregressive model performs the best in about 20 % when mainly the EE signal is flat. In rare cases, LSTM and CNN-LSTM outperform LSTM-LGBM and autoregressive models. The modeling approach with all features EE ML-0 performs the worst, while the approach with the top features performs the best. In most cases, EE ML-2 with less than 10 features outperforms EE ML-1 with top 10 features, while in rare cases, EE ML-1 slightly performs better. Thus, modeling with fewer of the most correlated parameters provides accurate prediction.

Examples of the comparison between the real and predicted values of EE for the ML-0, ML-1, and ML-2 approaches are illustrated in Fig. 6 for the personalized outputs of the ensemble LSTM-LGBM model. As a prediction, an average of 10 runs of LSTM models are shown for each case. Protocol A consisted of low-intensity activities; thus, the real and predicted EE values were in the narrow range of 0.8–2.2 kcal/min. The prediction accuracy MAPE for subjects S1, S4, and S5 is below 8 %. However, for the remaining subjects, only the mean of the EE signal is

correctly predicted, but the variations are not captured; therefore, the prediction accuracy is over 10 %. Protocol B, opposite to Protocol A, consisted of medium-intensity activities yielding EE values between the wide range of 0.8–5.5 kcal/min. The subject S5 had clear repeating patterns; thus, the best prediction accuracy was about 7 %. In the remaining cases, particularly for subject S3, the model prediction was over 13 % as fluctuations in the EE signal were not captured. Finally, Protocol C involved low-intensity activities with food intake, and thus the EE values lay in between the moderate range of 0.8–2.8 kcal/min. The best-predicted pattern was for the subjects S1-S2 where the MAPE was only 7–8 %. Only means were captured for subjects S3 and S6, although the prediction accuracy was 10 %. The worst performance was for subject S5 (MAPE was 15 %). Thus, from the comparison of the dynamic EE predictions across different protocols, we can outline that predictions below 10 % can be considered to be adequate, ideally, they should be down to 5 %.

To further highlight the improved performance of the personalized modeling approach, the performance of the ensemble model LSTM-LGBM was cross-tested by applying the features of other subjects to the personalized model developed for a specific subject. Fig. 7 provides the matrix of MAPE values (dimensionless, multiply the values by 100 to convert %) for Protocols A and B data. It could be noted that the model, for instance, developed for subject S1 using the S1-specific features, has the lowest MAPE compared to cases when non-S1-specific features are used. In rare cases, we can see the opposite, and it mainly relates to the signal of participants with very well-defined cyclic patterns that can be well captured by the LSTM-based models with a wide selection of features.

3.3. Discussion

The discussion will be focused on the comparison and improvement of personalized models over generalized model. We will assess if the models have integrated the four main effects that influence EE variation and investigate what are the most meaningful features. First of all, the MAPE values for all the models and subjects in the three protocols are summarized in Fig. 8 for the EE ML-1 approach considering the best 10 features. We can observe that the values for the generalized model are higher than the rest of the models (except the baseline in Protocols A and B). The genialized ensemble model captures only a global view of the EE signal as it uses a common ground of a few important features that might not be that important for everyone, it struggles to go into details in the prediction because each subject has its physiological characteristics. This reinforces the motivation of this work to demonstrate the better performance of personalized models for each subject and each protocol. For Protocols A and B, the ensemble LSTM-LGBM models led to errors under 15 % for all subjects (four subjects showed an error under 10 %), corresponding to RMSE below 0.3 kcal/min. For Protocol C, the overall MAPE is lower than the other protocols, and the baseline model also performs better; the lowest MAPE is under 12 %, corresponding to RMSE under 0.2 kcal/min. The LSTM-LGBM model is always the one that leads to the lowest error in Protocol C, which validates the expectation that it would outperform the simple LSTM and CNN-LSTM models. Indeed, the introduction of the LGBM model reduced the variability in the prediction compared to the CNN-LSTM model, so only the important changes were predicted. The error of the LSTM-LGBM model is 2–3 % less compared to the CNN-LSTM model and 5–8 % less compared to the simple LSTM.

To relate our prediction results with the results from the literature, the results of the generalized model can be compared with the ones identified in the literature and presented in Table 1. The best EE estimation was from the study of Sevil et al. [21], where the LSTM networks were also tested to estimate EE. In their study, 25 subjects underwent a sequence of physical activities for approximately 1 h, and sedentary state experiments lasted approximately 3 h. Some subjects repeated certain physical activities. They used this database of 71 h of data for training and conducted a separate new experiment of 110 min with one of the subjects performing physical activities with resting and some everyday living activities to evaluate the EE estimation accuracy. With this one experiment, a MAPE of 12.1 % was achieved. In our work, we achieved comparable results using the generalized ensemble model with CNN-LSTM and LGBM and reached a MAPE ranging from 8.8 % to 47.7 %, with a mean error of 14 % for Protocol A, 17.6 % for Protocol B and 13.6 % for Protocol C. The prediction accuracy of the personalized models is much better, reaching the MAPE down to 5 % in some cases, which corresponds to very good prediction results.

As the main intention of this work was to consider more effects on the energy expenditure EE prediction apart from the physical activity, the personalized approach to modeling allowed to integrate the effects of the ambient temperature, food-induced thermogenesis, and the sex effect. The following are the main observations from our results:

1. ***Physical activity effect***: For low-intensity activities (Protocols A and C), the predictions have an error down to 5 %, which is quite small. In this context, accelerometry data from the chest and thigh have a high predictive power, as it is the variable that catches movements of the lower and the upper body during sitting and standing. Accelerometry is still important in higher intensities, but heart rate becomes the main parameter. Subjects going through physical efforts would have an elevated heart rate.
2. ***Sex effect***: Differences in EE prediction can be noted according to the subject's sex. For example, in Protocol C, the mean of the EE signal is considerably lower in women than men. Fig. 6 shows poorer prediction in women compared to men. This could be due to the feature importance, as in women, skin temperature variables are more important than in men, and so accelerometry data is less represented. Therefore, the actual reason for the difference in the magnitude of the EE values might come from the physiological difference between males and females.
3. ***Ambient temperature effect***: For the change in ambient temperature, we can see the effect on the EE signal very clearly for S6 in Protocol A, where the signal seems to increase slightly during the session, probably because the subject is cold. The effect seems to be captured by the personalized models, as we see a slight increase in the prediction in Fig. 6 as well. The important features for this subject are the skin temperature at the left and right instep and the skin temperature at the wrist, so it seems that skin temperature at the body's extremities is important to detect ambient temperature variations. For Protocol B, the effect of the change in ambient temperature is not represented in the EE signal because the physical activity suppresses it. In the generalized model, the constant increase in EE signal for S6 in Protocol A is not captured by the prediction, so the model fails to represent it; also, some temperature-related variables, such as the instep skin temperature, are not present in the generalized model.
4. ***Food intake effect***: The effect of food intake on the EE signal is well represented in S1, S2, and S6 in Protocol C. The variations of the EE signal are more accurately predicted in men than for women, where the predictions do not follow the small variations. From the feature importance, we can notice that the heart rate is a very important parameter. Indeed, some studies demonstrated that meals have an impact on heart rate. Chang et al. [32] showed that, directly after the meal, the heart rate increases significantly from baseline and decreases significantly from 40 min to 120 min after the meal. The study suggests that heart rate variability reflects cardiac and also abdominovagal activity, but not equivalently. This implies that a change in heart rate due to food intake could be expected in some subjects. Also, for females, skin temperature is more important than in males. This could also be an effect of food intake. Martinez et al. [33] reported that the overall, mean, proximal, and supraclavicular skin temperature significantly increased after the meal intake and that there is a postprandial peripheral vasoconstriction right after the meal intake and over the first hour and a peripheral vasodilatation

during the second and third hour. Women have a higher increase in all skin temperature parameters than men. This could explain why skin temperature is important mostly for women but also sometimes for males in the experimental protocol with food intake.

4. Conclusions

This paper aimed to develop a personalized dynamic human body energy expenditure prediction model using the LSTM network. In most of the prior similar works, the researchers considered only physical activity to develop the EE prediction model. This work proposed to achieve a better prediction of EE from secondary measurements by considering the effect of physical activity, changes in ambient temperature, and food intake. To these objectives, experiments with multiple protocols were conducted with 6 test subjects (3 males and 3 females) across 3 different protocols. The protocols were designed to explore the variation in energy expenditure with respect to varying air temperature, activity levels, and food intake. The analysis showed that the EE could be quite accurately predicted during low-intensity activities by measuring skin temperature and body accelerometry data. The skin temperature measurements with a good correlation with the target variable were mostly from the chest, insteps, calves, arms, shins, and forehead. In the case of the accelerometry data, the same was obtained from the upper chest, right thigh, and right hand. For medium-intensity activities, the heart rate and accelerometry data provided the highest F-values with respect to EE.

For the development of the prediction model, three different types of feature selection were adopted. Since the dynamics of EE in the human body differs between people, the models that were developed varied in terms of the set of features that were selected for each. Based on the selected feature sets, the LSTM prediction models were developed. The results show that the ensemble LSTM-LGBM models provide a good level of prediction accuracy during both low and medium-intensity activities, with the MAPE mostly lying in the range of 5–15 %. The approach presented in the paper considers the temperature effect in addition to the variation in activity levels. It is an improvement on the previous studies that considered only accelerometry and heart rate data for developing the human energy expenditure prediction model. Considering the results with the generalized models, we conclude that personalization was crucial, as the feature set for the prediction models differed completely between the protocols and subjects and that they have succeeded in catching the effects of physical activity, ambient temperature, and food intake. The personalized modeling approach with LSTM has shown the potential to improve prediction accuracy compared to the pre-existing models. Moreover, the use of ensemble models combining LSTM with the Gradient Boosting algorithm shows a novelty in the approach and has proven to decrease the prediction error by 8 % compared to a simple LSTM model. One of the main limitations of the LSTM-based models is that they focus on detecting patterns and work well for signals with specific repetitive features, as ated in this in Thus, the model might not be suitable for predicting short-term events, but focusing on the scale of the days, when people typically have repetitive routines day-by-day, would make using the LSTM model suitable. Another practical limitation can be linked to the time required to run the models. The LSTM-based models are stochastic by nature and need training the models repetitively multiple times (10 runs in our work). In addition, the best-performing model was the ensemble model, which is relatively heavy-weight. Therefore, future works should focus on developing a more lightweight configuration of the models.

CRediT authorship contribution statement

Victoria M. Perez Cortes: Methodology, Formal analysis, Data curation, Visualization, Writing – original draft. **Arnab Chatterjee:** Methodology, Formal analysis, Data curation, Writing – review & editing. **Dolaana Khovalyg:** Conceptualization, Investigation, Resources, Supervision, Funding acquisition, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to thank Prof. Abdul Dulloo, Prof. Jean-Pierre Montani, Prof. Yves Schutz, and Dr. Yann Ravussin from the University of Fribourg for their suggestions on the design of experiments and their help with the use of the indirect calorimeter. A special acknowledgment to Isabelle Scerri and Jiyoung Kwak for their assistance in conducting the experiments. Finally, we thank all participants that took part in our study.

Appendix A: Tuned hyperparameters for LSTM-based ML models

(See Table A1 and See Table A2).

Appendix B: Data variability examples

The examples of variation in the skin temperature and accelerometery measurements at various body parts in all the protocols for the male subjects S2 and the female S4 are presented in Figs. A1, A2 The plots are divided into 3 parts illustrating the measurements from three different protocols. For protocols A and B, the differences in the skin temperature during sessions 1 and 3 at ambient temperatures of 24 °C and 18 °C are shown as violin plots. Similarly, for Protocol C, the violin plots illustrate the differences in the skin temperatures at different body parts during sessions 2 and 3, before and after the mid-day meal. Going from left to right in the plot signifies going from core body temperature measurements to the extremities. It is quite evident the fact that in protocols A and B, the difference in skin temperatures at 18 and 24 °C is much higher (about 5–10 °C) in the extremities of the body, such as the fingers, insteps, wrist, and shins in comparison to the core body parts (1–3 °C) such as the chest, abdomen, and hip. In the case of protocol C, the magnitude of this difference is much lower between before and after lunch measurements since the ambient temperature was kept more or

Table A1
Fine-tuned parameters for training of personalized LSTM and CNN-LSTM models.

Protocol ID	Subject ID	Window Size	Epochs	Batch size	Neurons
A	1	250	50	100	100
	2	250	100	64	100
	3	200	50	64	50
	4	150	100	64	100
	5	200	50	64	100
	6	200	50	64	50
B	1	180	50	100	50
	2	250	100	64	50
	3	200	50	100	50
	4	250	100	64	100
	5	200	50	128	50
	6	200	50	64	50
C	1	250	100	100	100
	2	200	100	100	50
	3	200	50	128	100
	4	150	50	64	100
	5	200	50	64	100
	6	200	50	128	50

Table A2
Fine-tuned parameters for training of the personalized LSTM-LGBM model.

Protocol ID	Subject ID	Model	Colsample by tree	Learning rate	Max depth	Min child weight	Subsample	n-est
A		EE ML-0	0.7	0.15	6	7	0.921631171	100
	1	EE ML-1	0.3	0.05	5	6	0.852823881	100
		EE ML-2	0.3	0.3	5	2	0.843762985	100
		EE ML-0	0.3	0.25	6	6	0.910062358	100
	2	EE ML-1	0.5	0.25	7	2	0.851981787	100
		EE ML-2	0.3	0.3	13	1	0.932981864	100
		EE ML-0	0.7	0.25	15	5	0.802312547	100
	3	EE ML-1	0.3	0.1	8	4	0.819527643	100
		EE ML-1	0.4	0.1	5	4	0.887204318	100
		EE ML-0	0.5	0.05	6	3	0.992031798	100
	4	EE ML-1	0.3	0.2	15	6	0.825367815	100
		EE ML-2	0.6	0.2	5	1	0.879214255	100
		EE ML-0	0.4	0.05	7	5	0.950560582	100
	5	EE ML-1	0.4	0.1	5	2	0.898274478	100
		EE ML-2	0.4	0.2	6	2	0.899410153	100
		EE ML-0	0.4	0.05	7	2	0.905112462	100
	6	EE ML-1	0.4	0.25	6	1	0.921165118	100
		EE ML-2	0.5	0.3	5	1	0.940259347	100
B		EE ML-0	0.7	0.3	6	5	0.882690167	100
	1	EE ML-1	0.6	0.25	5	4	0.92787486	100
		EE ML-2	0.6	0.3	14	3	0.895616306	100
		EE ML-0	0.3	0.1	15	3	0.914959555	100
	2	EE ML-1	0.4	0.3	5	4	0.882874568	100
		EE ML-2	0.4	0.3	10	6	0.894935462	100
		EE ML-0	0.6	0.1	9	1	0.95373152	100
	3	EE ML-1	0.7	0.1	6	2	0.843179479	100
		EE ML-2	0.5	0.2	6	2	0.837384932	100
		EE ML-0	0.4	0.15	7	8	0.875352154	100
	4	EE ML-1	0.3	0.3	8	1	0.845634082	100
		EE ML-2	0.7	0.3	12	7	0.801412667	100
		EE ML-0	0.7	0.15	12	7	0.854741552	100
	5	EE ML-1	0.4	0.15	11	3	0.920435347	100
		EE ML-2	0.4	0.15	9	4	0.926653577	100
		EE ML-0	0.4	0.05	7	6	0.978610101	100
	6	EE ML-1	0.6	0.05	6	7	0.925882483	100
		EE ML-2	0.7	0.25	6	2	0.925724279	100
C		EE ML-0	0.3	0.05	7	6	0.996316451	100
	1	EE ML-1	0.7	0.05	5	6	0.938052335	100
		EE ML-2	0.4	0.05	6	6	0.922475163	100
		EE ML-0	0.3	0.25	7	6	0.934654673	100
	2	EE ML-1	0.4	0.1	5	2	0.959373106	100
		EE ML-2	0.3	0.3	5	5	0.867055375	100
		EE ML-0	0.3	0.15	15	3	0.823409868	100
	3	EE ML-1	0.6	0.15	12	7	0.817610498	100
		EE ML-2	0.3	0.2	10	7	0.801326352	100
		EE ML-0	0.7	0.3	7	5	0.857275938	100
	4	EE ML-1	0.4	0.3	6	5	0.855054402	100
		EE ML-2	0.7	0.3	10	7	0.820586621	100
		EE ML-0	0.4	0.2	6	3	0.972447915	100
	5	EE ML-1	0.3	0.2	9	4	0.816523015	100
		EE ML-2	0.5	0.3	10	6	0.826734973	100
		EE ML-0	0.5	0.25	6	4	0.911038948	100
	6	EE ML-1	0.7	0.1	11	6	0.922457379	100
		EE ML-2	0.7	0.2	7	6	0.865824121	100

less constant in this protocol. Comparing male and female subjects, the temperature of extremities (e.g., fingers, instep) tended to be lower in the female S4, especially in protocol C.

To demonstrate variability in the accelerometery measurements at various body parts in Fig. A2, violin plots carry different information for the left and the right parts. For protocols A and B, the left half of the violin plot shows the variation of the activity levels along the ×, y, and z axis for the ambient temperature 18 °C, while the right half shows for 24 °C. In Protocol A, the violin plots show the accelerometry data when the subjects were either sitting or standing, while for Protocol B, the traces show the acceleration value along the three axes for when the subject was either performing a low-intensity (0 W) cycling or high- intensity cycling (40 W). Similarly, for Protocol C, the violin plots illustrate the differences in the activity levels at different body parts before and after lunch for either sitting or standing activity. It is quite evident that the magnitude of the activity levels in Protocol B is higher than that in Protocols A and C for both subjects. Comparing male and female subjects, there is not much difference in the accelerometry data variation as they were standardized.

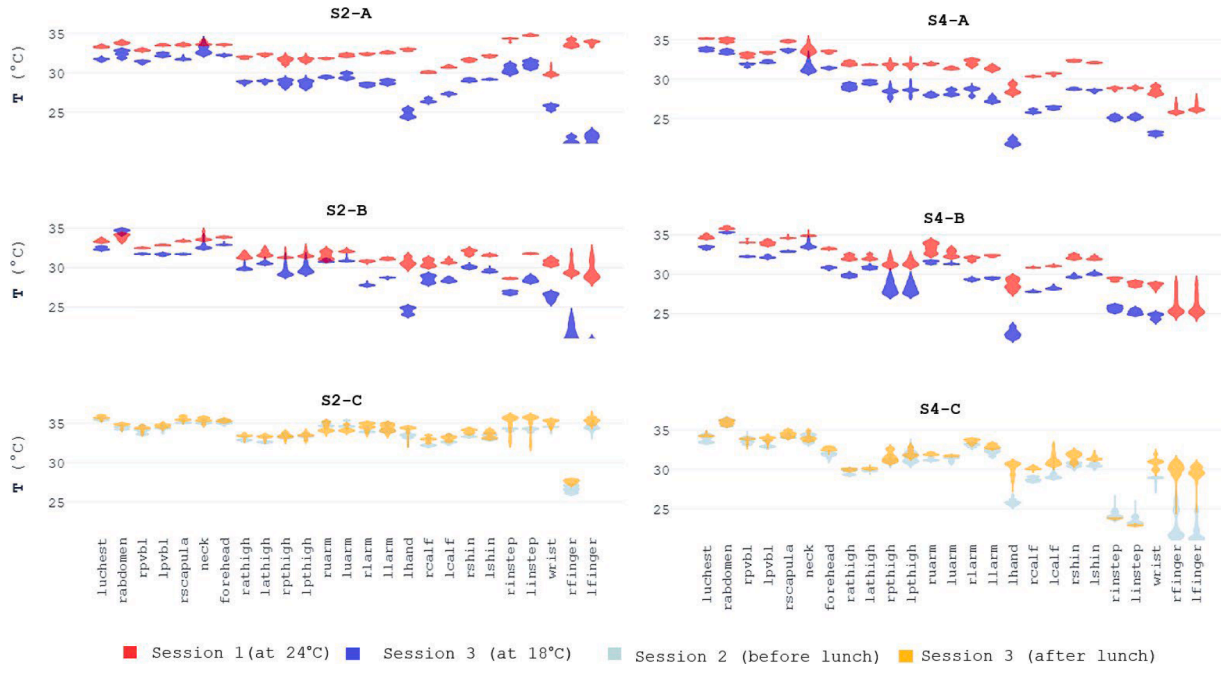


Fig. A1. Comparison of the skin temperatures distribution for subjects S2 (male) and S4 (female) for selected parts of three protocols (Protocol A and B: 18 °C vs. 24 °C, Protocol C: before vs. after lunch).

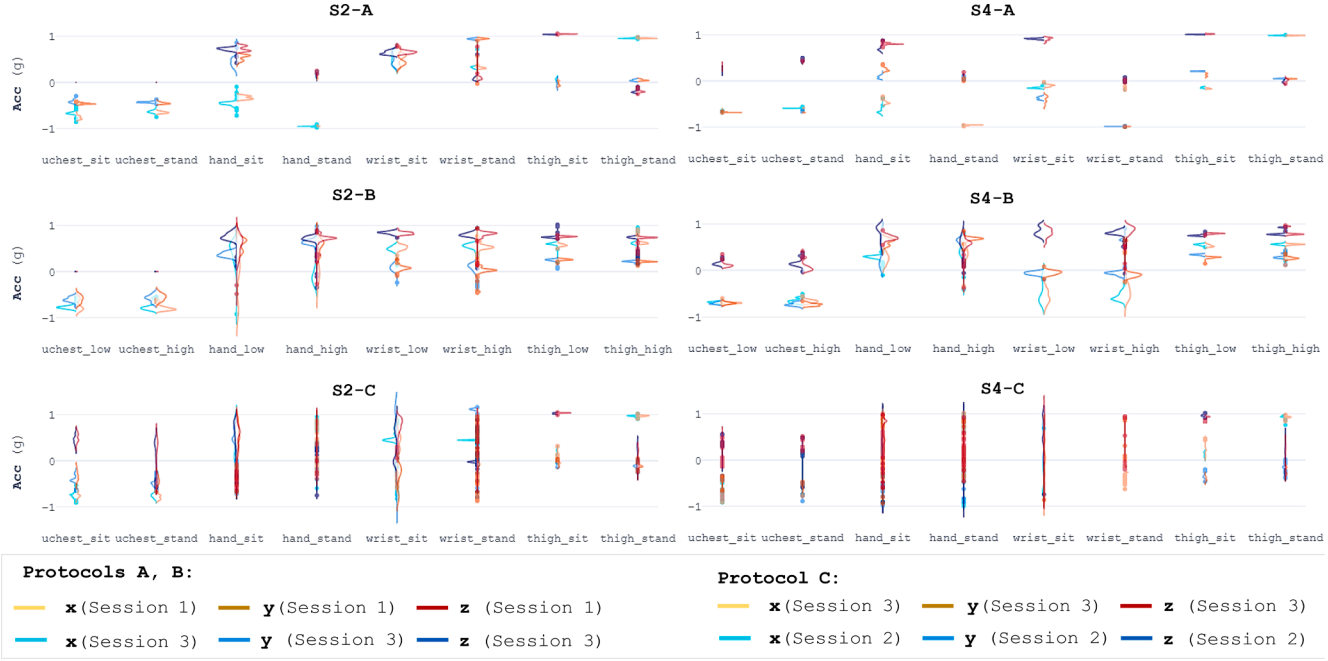


Fig. A2. Comparison of the accelerometery data distribution, per axes, for subjects S2 (male) and S4 (female) for selected parts of three protocols (Protocol A: sit vs. stand, Protocol B: 0 W (low) vs. 40 W (high) cycling efforts, Protocol C: before vs. after lunch).

References

[1] L. Ruddick-Collins, A. Flanagan, J. Johnston, P. Morgan, A. Johnstone, Circadian Rhythms in resting metabolic rate account for apparent daily rhythms in the thermic effect of food, J. Clin. Endocrinol. Metabol. 107 (2) (2022) 708–715.

[2] J. Levine, Non-exercise activity thermogenesis (NEAT), Best Pract. Res. Clin. Endocr. Metabol. 16 (4) (2002) 679–702.

[3] A. Dulloo, J. Miles-Chan, J. Montani, Y. Schutz, Isometric thermogenesis at rest and during movement: a neglected variable in energy expenditure and obesity predisposition, Obes. Rev. 18 (S1) (2017) 56–64.

[4] N. Chung, M. Park, J. Kim, H. Park, H. Hwang, C. Lee, J. Han, J. So, J. Park, K. Lim, Non-exercise activity thermogenesis (NEAT): a component of total daily energy expenditure, J. Exer. Nutr. Biochem. 22 (2) (2018) 023–030.

[5] A. Sellers, D. Khovalyg, G. Plasqui, W. Lichtenbelt, High daily energy expenditure of Tuvan nomadic pastoralists living in an extreme cold environment, Sci. Rep. 12 (2022).

[6] H. Pallubinsky, L. Schellen, W. van Marken Lichtenbelt, Exploring the human thermoneutral zone – A dynamic approach, J. Therm. Biol. 79 (2019) 199–208.

[7] G.P. Kenny, S.R. Notley, D. Gagnon, Direct calorimetry: a brief historical review of its use in the study of human metabolism and thermoregulation, Eur. J. Appl. Physiol. 117 (9) (2017) 1765–1785.

[8] H. Mtaweh, L. Tuira, A. Floh, C. Parshuram, Indirect calorimetry: History, technology, and application, Front. Pediatr. 6 (257) (2018) pp.

[9] E. Zhou, H. Hu, Human motion tracking for rehabilitation—A survey, Biomed. Signal Process. Control 3 (1) (2008) 1–18.

[10] N. Hegde, M. Bries, E. Sazonov, A comparative review of footwear-based wearable systems, Electronics 5 (3) (2016) 48.

[11] R. O'Driscoll, J. Turicchi, K. Beaulieu, S. Scott, J. Matu, K. Deighton, G. Finlayson, J. Stubbs, How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies, British J. Sport Med. 54 (2020) 332–340.

[12] D. Hendelman, K. Miller, C. Baggett, E. Debold, P. Freedson, Validity of accelerometry for the assessment of moderate intensity physical activity in the field, Med. Sci. Sports Exer. 32 (9) (2000) 442–449.

[13] G. Welk, S. Blair, K. Wood, S. Jones, R. Thompson, A comparative evaluation of three accelerometry-based physical activity monitors, Med. Sci. Sports Exer. 32 (9) (2000) 489–497.

[14] A. Swartz, S. Strath, D. Bassett, W. O'Brien, G. King, B. Ainsworth, Estimation of energy expenditure using CSA accelerometers at hip and wrist sites, Med. Sci. Sports Exer. (2000) 450–456.

[15] D. Spierer, M. Hagins, A. Rundle, E. Pappas, A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging, Eur. J. Appl. Physiol. 111 (4) (2011) 659–667.

[16] K. Dannecker, N. Sazonova, E. Melanson, E. Sazonov, R. Browning, A comparison of energy expenditure estimation of several physical activity monitors, Med. Sci. Sports Exer. 45 (11) (2013) 2105–2112.

[17] J. Zhu, A. Pande, P. Mohapatra and J. J. Han, Using Deep Learning for Energy Expenditure Estimation with wearable sensors, in: 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), Boston, MA, USA, 2015, pp. 501-506.

[18] B. Cvetkovic, R. Milic, M. Lustrek, Estimating energy expenditure with multiple models using different wearable sensors, IEEE J. Biomed. Health Informat. 20 (4) (2016) 1081–1087.

[19] A. Montoye, J. Pivarnik, L. Mudd, S. Biswas, K. Pfeiffer, Wrist-independent energy expenditure prediction models from raw accelerometer data, Physiol. Measur. 37 (2016) 1770–1784.

[20] R. O'Driscoll, J. Turicchi, M. Hopkins, G.W. Horgan, G. Finlayson, J.R. Stubbs, Improving energy expenditure estimates from wearable devices: A machine learning approach, J Sports Sci. 38 (13) (Jul; 2020,) 1496–1505.

[21] M. Sevil, M. Rashid, Z. Maloney, I. Hajizadeh, Determining physical activity characteristics from wristband data for use in automated insulin delivery systems, IEEE Sens J 20 (21) (2020) 12859–12870.

[22] M. Mohamed, S. Refaat Shady, et al., A novel stacked generalization ensemble-based hybridLGBM-XGB-MLP model for Short-Term Forecasting, Energy 214 (2021), 118874.

[23] Qian Wei, GU, Chunlei, ZHU, Congxi et al., Short-Term Load Forecasting Based on Multi-model Fusion of CNN-LSTM-LGBM, in: 2021 International Conference on Power Systerm Technology (POWERCON), IEEE, 2021, p. 934–939.

[24] P.B. Weerakody, K.W. Wong, G. Wang, W. Ela, A review of irregular time series data handling with gated recurrent neural networks, Neurocomputing 441 (2021) 161–178.

[25] N. Pai and V. Ilango, LSTM neural network model with feature selection for financial time series prediction, in: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 672-677.

[26] D. Khovalyg, Y. Ravussin, Inter-individual variability of human thermoregulation: towards personalized ergonomics of the indoor thermal environment, Obesity 30 (7) (2022) 234–254.

[27] A. Heidari, D. Khovalyg, DeepValve: Development and experimental testing of a Reinforcement Learning control framework for occupant-centric heating in offices, Eng. Applicat. Artif. Intel. 123 (2023), 106310.

[28] M. Rahiminejad, D. Khovalyg, Experimental study of the hydrodynamic and thermal performances of ventilated wall structures, Build. Environ. 233 (2023), 110114.

[29] D. Khovalyg and J. Kwak, Dynamics of metabolic rate in male individuals due to the meal and regular office activities, in: CLIMA 2022 The 14th REHVA HVAC World Congress Proceedings, 2022.

[30] J. Weir, New methods for calculating metabolic rate with special reference to protein metabolism, The Journal of Physiology 109 (1–2) (1949) 1–9.

[31] Source Code - Github Repository, 2023, https://zenodo.org/record/7584890.

[32] C.S. Chang, C.W. Ko, H.C. Lien, M.C. Chou, Varying postprandial abdominovagal and cardiovagal activity in normal subjects, Neurogastroenterol Motil 22 (5) (2010) 546–551, e119.

[33] B. Martinez-Tellez, L. Ortiz-Alvarez, Skin temperature response to a liquid meal intake is different in men than in women, Clin Nutr. 38 (3) (2019) 1339–1347.