Building and Environment 245 (2023) 110850

Contents lists available at ScienceDirect

Building and Environment

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

Toward contactless human thermal monitoring: A framework for Machine Learning-based human thermo-physiology modeling augmented with computer vision

Mohamad Rida [a], Mohamed Abdelfattah [b], Alexandre Alahi [b], Dolaana Khovalyg [a,*]

[a] Laboratory of Integrated Comfort Engineering (ICE), École polytechnique fédérale de Lausanne (EPFL), Fribourg, Switzerland
[b] Visual Intelligence for Transportation (VITA), École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Keywords:
Thermal comfort
Human thermo-physiology modeling
Non-intrusive sensing
Artificial intelligence
Deep learning
Computer vision

ABSTRACT

The transition towards a human-centered indoor climate is beneficial from occupants' thermal comfort and from an energy reduction perspective. However, achieving this goal requires the knowledge of the thermal state of individuals at the level of body parts. Many current solutions rely on intrusive wearable technologies, which require physical access to individuals facing limitations in scalability. Personalizing the indoor environment demands increased sensing at individual levels presenting challenges in terms of data collection and ensuring privacy protection. To address this challenge, this paper introduces a novel approach to non-intrusive personalized humans thermal sensing that can acquire personal data while minimizing the amount of sensing required. The method investigates multi-modal sensing solutions based on IR and RGB images, and it includes the development of a Machine Learning-based Human Thermo-Physiology Model (ML-HTPM). With the help of computer vision, features important for thermal comfort such as activity level, clothing insulation, posture, age, and sex can be extracted from an RGB image sequence using models such as the SlowFast network, YOLOv 7, while limited skin temperatures can be extracted from an IR image using OpenPifPaf for body parts detection. The developed ML-HTPM is based on data generated from an open-source JOS3 model after applying a prediction model based on Long Short-Term Memory (LSTM). The results showed that a human thermo-physiology model using machine learning can be trained, showing an RMSE of less than 0.5 °C in most of the local skin temperatures.

1. Introduction

Rapid urbanization and the fact that people spend almost 90% of their time indoors makes indoor environmental quality (IEQ) accountable in new and existing buildings for assuring the well-being of the occupants [1]. IEQ is characterized by environmental categories such as thermal, air quality, lighting, and acoustics. While each category is important for the comfort and well-being of occupants, thermal (dis)comfort is the most familiar and easily recognizable by occupants. Therefore, the indoor temperatures are typically set in buildings with the aim of providing *thermal comfort* to occupants by keeping their sensation around *thermal neutrality* (the state when a human body primarily maintains its core body temperature with minimal metabolic regulation) [2,3]. However, the practice of setting the indoor temperature at a narrow range has resulted in almost 40% of operational energy use in buildings [4,5]. Surprisingly, as evidenced from multiple field studies, e.g., [6,7], the buildings that are supposedly designed according to the standardized requirements do not necessarily provide a satisfactory experience to their occupants primarily because of human diversity. The current methods of setting the indoor climate consider *an average person* [8], but *"one size does not fit all"* [9] and individual differences in thermal sensation are well documented in the literature [10,11]. Therefore, in quest of improving the well-being of occupants and limiting energy use in buildings, there is a need to advance methods to detect the thermal sensation of individuals and potentially incorporate them into the control of the climatization system of the building.

As humans are endotherms, the core body temperature (T_{core}) in healthy people is rather stable even for a wide range of ambient temperatures and during exercises, excluding scenarios of hypo- and hyperthermia [12,13]. The gradient between the skin temperature of the core body part such as a chest ($T_{skin,chest}$) and other body parts ($T_{skin,i}$), particularly of extremities, drives thermal sensation of individuals [14]. Zhang et al. [15] and Choi and Loftness [16] conducted

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

https://doi.org/10.1016/j.buildenv.2023.110850
Received 19 June 2023; Received in revised form 3 September 2023; Accepted 18 September 2023
Available online 26 September 2023

studies on the factors affecting the prediction of thermal sensation. They found that skin temperature exhibited the highest correlation among physiological parameters, capturing the combined influence of environmental factors, body thermoregulation, and individual characteristics. According to Bulcao et al. [17], skin temperature is the primary determinant of thermal comfort in humans, whereas both skin temperature and core temperature variations contribute to physiological reactions related to thermal changes [18,19]. Thus, the overall thermal sensation is a function of the local sensation of individual body parts [14], and knowing the local skin temperature at different body parts ($T_{skin,i}$) is important to advance the human-centric thermal environment and personalized conditioning systems that can aid the reduction of energy use in buildings [20–22].

1.1. Human thermal state modeling

Mathematical models of human thermoregulation called *human thermo-physiology models* (HTPMs) can estimate the change in core temperature T_{core} and local skin temperature $T_{skin,i}$ at steady and transient thermal environments [23]; thus, serve as a tool to predict local thermal comfort of people. Most of the detailed multi-node HTPMs are based on Stolwijk's model [24,25], examples are Fiala's model [26], Tanabe's model (JOS1-JOS3) [23], the Berkeley Comfort Model [27], the AUB model [28], and ThermoSEM [29,30]. All models are based on the energy balance between the environment and the human body; thus, they include heat transfer between the skin layer and the environment in addition to conduction between the different skin layers and convection due to the blood circulation between layers and body parts [23,26,28,30]. The typical inputs used in HTPMs are *environmental parameters* (local air temperature T_{air}, mean radiant temperature MRT, air speed V_{air}, relative humidity RH) and *personal parameters* (activity level Act, local clothing insulation I_{Cl}, and sometimes body characteristics). Environmental parameters should be input at each body part, which is typically challenging to determine; nevertheless, there has yet to be an HTPM that could be executed without input information about the environment.

Models ThermoSEM and JOS3 can use individual body parameters as input and be re-scaled to fit an individual, thus, they are capable of predicting personalized local thermal responses. JOS3, a description of which is detailed in Appendix A, is an open-access model [23] and, per a validation study by Rida et al. [31], it has better accuracy in predicting skin and core temperatures compared to ThermoSEM. JOS3 model showed a root-mean-squared-error RMSE of 0.3 °C for core temperature and 0.9 °C for mean skin temperature; therefore, this model could be used to determine the local temperatures of the human body, however, accurate environmental data and personal parameters such as *clothing* and *metabolic rate* must be input. Since metabolism is quite individual [32], the input of additional individual characteristics (*e.g.*, height, weight, age, and gender) are essential. Thus, the model JOS3 is capable of detailing the interaction of humans with their environment, but it requires multiple environmental and personal sensing and the knowledge of individual parameters.

1.2. Human comfort, environment, and personalized features detection

Currently, the most direct way of knowing the local thermal sensation of people is surveying, which is limited by the participation rate [33]. Typical monitoring of environmental parameters at the room level using a restricted number of sensors is difficult to match with the actual thermal sensation of humans since they are usually attached to building surfaces away from people [34]. Detailed measurements of physiological changes in the human body caused by the environment (*i.e.*, vasodilation, vasoconstriction) can precisely inform regarding the local thermal sensation of individuals; however, it is difficult to perform such measurements outside the lab. With the advancement of wearable devices, it has become possible to monitor certain physiological parameters of humans (*i.e.,* HR, wrist skin temperature, etc.) in the actual environment and to correlate them with individuals' thermal sensation [35,36]. However, the measurements using wearables are still relatively imprecise, and physiological input to wearable devices is limited [34]. Alternatively, there are emerging *data-driven methods* combining objective environmental measurements with subjective feedback [22,37]; but they require large dataset collection using *intrusive* or *semi-intrusive* approaches [38]. Data-intensive methods contradict the dislike of occupants to be equipped with sensors and respond to multiple surveys. Therefore, minimizing the number of data sources and maximizing the data extraction from limited sources is a challenge to overcome. With this respect, non-intrusive measurements using cameras augmented with *computer vision methods* are gaining more attention due to the recent advances in Deep Learning.

Multiple researchers attempted to develop machine learning models to predict thermal comfort and showed that skin temperatures were important predictors. Yu et al. [39] tested different machine learning algorithms (e.g., Support Vector Machine, Decision Tree, K-Nearest Neighbors) to predict thermal comfort, and head and hand skin temperatures were the main contributors to the model. Other studies, such as [40,41], focused on correlating some areas of the face temperature to thermal comfort. Lyu et al. [40] used the classification Random Forest algorithm to model thermal sensation based on facial skin temperature extracted from an IR image with the help of the OpenCV library to detect the face. They determined that the frontal view of the face had the best results, followed by the lateral view of the face. Li et al. [41] studied the impact of the temperature collection zones that can accurately reflect the thermal sensation by applying the Decision Forest algorithm in addition to the Haar Cascade algorithm for face detection. Ghahramani et al. [42] used hidden Markov model, a continuous learning method, to capture personal thermal comfort using IR thermography of the human face. Jazizadeh and Jung [43] used RGB video camera combined with Eulerian Video Magnification algorithm to infer personalized blood perfusion state through tracking of facial skin color variations. Moreover, Li et al. [44] also examined the use of an IR thermal camera network to identify skin temperature features and predict occupants' thermal preferences at variable angles and distances. Based on their experiment, they reported that the variations in skin temperature correspond to the comfort states reported by the subjects. Studies by Aryal and Becerik-Gerber [45,46] compared different sensing methods including the use of an IR camera FLIR Lepton for overall thermal comfort prediction using multiple algorithms (Random Forest, KNN, SVM, and Decision Trees). Comfort prediction using only a thermal camera was not the best one, and the lower accuracy was attributed to the high levels of noise in the data. Thus, the IR camera quality influences the comfort prediction accuracy.

Recently, deep convolution-based models [47,48] have been proven significantly efficient in predicting personalized features such as the pose, age, gender, activity, clothing of humans solely based on input RGB images. Multiple methods [49,50] have been proposed to extract human poses from input RGB images. Accurate prediction of age and gender has been achieved by training simple Multilayer Perceptron (MLPs) [51] on publicly available datasets such as [52,53] and it achieves the best age and gender classification accuracies of 60% and 91% respectively on the Audience dataset. Additionally, RGB images can be used to accurately detect the types of clothing items of humans in the image. The introduction of large-scale fashion datasets [54,55] makes it possible to train generic state-of-the-art object detection models [56,57], paving the way for automatic, non-intrusive detection of clothing items of humans in the field of view of a camera. Specifically, the model YOLOv7 [56] on DeepFashion dataset [55] is popular for clothes detection as it is lightweight and has superior Average Precision (AP) of 56.8% on MS COCO dataset [58] and state-of-the-art performance on multiple other object detection datasets [59,60]. Finally, recent deep learning models [61,62] have been developed to detect and classify human actions by capturing spatial and temporal information

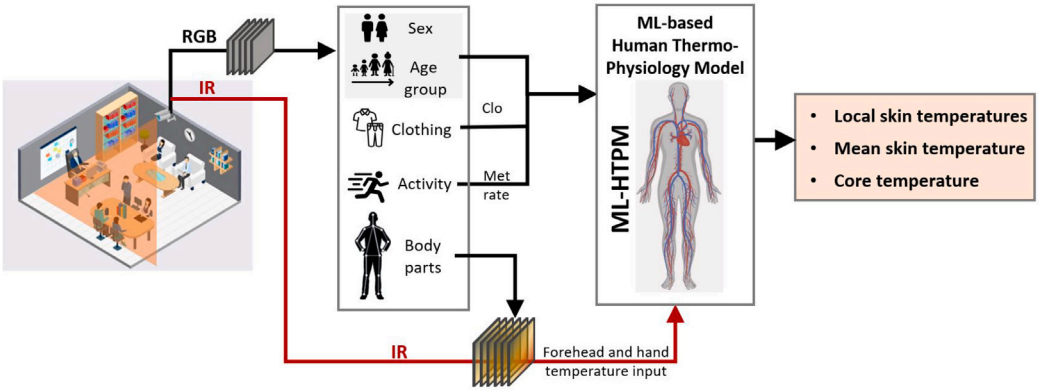

Fig. 1. Schematic of the proposed non-intrusive sensing solution involving IR+RGB imaging for personalized thermal sensing.

from input RGB videos. Among the most successful models is the SlowFast [63] network which relies on 3D convolutions for accurate action detection. SlowFast has been proven considerably effective in detecting actions from generic in-the-wild videos, achieving top-1 accuracy of 79.8% on the challenging Kinetics-400 dataset [64]. The integration of such information about the pose, activity, age, gender, and clothing insulation constitutes a viable input to machine learning regression models to accurately predict body skin temperatures and, hence, human thermal comfort.

1.3. Objectives of the current work

This work tests the capability of the multi-modal non-intrusive sensing solution involving IR and RGB imaging to extract thermal comfort-related features of different people located in the field of view of the camera. To this aim, we exploit the human thermo-physiology model JOS3 and develop a new Machine Learning model, referred to as ML-HTPM, that uses as input a restricted set of individual parameters that are detectable with our proposed non-intrusive camera-based sensing method. Thus, the proposed framework contributes to the advancement of scalable and cost-effective personalized comfort monitoring solutions. Contactless monitoring of the *local* and *overall* comfort of individuals can be used for (i) better understanding if the building meets comfort criteria, and (ii) for personalized climate control. Instead of using the temperature input from the conventional thermostats attached to the walls, away from people, the camera-based solution could provide input in terms of the actual local and overall thermal state of people to the controls of both personalized and centralized climatic systems.

2. Methodology

Our method of detecting human thermal comfort-related parameters and the prediction of individual thermal state non-intrusively using RGB and IR cameras features the following steps as shown in Fig. 1:

- Extraction of individual thermal comfort-related parameters (sex, age group, clothing type, and activity type) in real-time by deploying advanced computer vision techniques using live-streamed RGB images.
- Extraction of local skin temperatures of body parts such as head and hand from IR images supported by the body key points detection from RGB images. Two body parts were chosen to capture distal-proximal skin temperature gradient. The accuracy of the IR camera was supplementary evaluated by comparing IR measurements with contact skin temperature sensors in Appendix B.
- Development of the Machine Leanring-based human thermo-physiology model (ML-HTPM) that uses head and hand skin temperature and personal factors that can be extracted from RGB images as inputs to predict an individual's thermal state.

2.1. Computer vision for personalized features detection

We used several computer vision models to extract personalized features of human subjects as illustrated in Fig. 2. For age and gender detection, we employed RetinaFace [65] for face detection, ArcFace [66] for face feature extraction, and a Multilayer Perceptron (MLP) [51] for age and gender prediction. Action detection was achieved by leveraging the SlowFast [63] network with a 3D-ResNet backbone. All actions were classified according to activity types listed in standards ASRHAE 55 to refer to their corresponding estimated metabolic rate value in *met* as described in Appendix D (Table D.3). Further, we detected the types of clothing by training YOLOv7 [56] object detection model from scratch on the DeepFashion2 [67] dataset. Different clothing types detected were matched to the classification of clothing in standard ISO7730 [68] to be able to match with the standard clothing insulation values (Appendix C, Fig. C.19). Finally, for posture and body parts detection, the machine learning library OpenPifPaf [69] was used. It provides detailed keypoint annotations for the face, hands, and feet, enabling accurate pose estimation from RGB images. By using synchronized RGB and IR image streaming, the model can be used to locate the coordinates of the head and hands in RGB images and extract the temperatures of corresponding body parts from the IR thermal image as demonstrated in Fig. 3. Many modern IR cameras are equipped with a synchronized RGB camera, allowing for the simultaneous capture of RGB and IR images. This approach allows for the extraction of the body temperature, bridging the visual representation of the body with thermal patterns. Overall, these models contribute to non-invasive thermal sensation prediction by providing information on factors such as age, gender, actions, clothing, and body posture in relation to thermal comfort. A detailed description of each model is provided in Appendix C.

2.2. ML-HTPM development

The traditional use of the HTPM requires environmental parameter inputs like air temperature (T_{air}), mean radaint temperature (MRT), relative humidity (RH), and air speed (V_{air}). In addition, it requires personal parameters like activity (Act), clothing insulation (Clo) and body composition to predict the person's local skin temperature T_{skin} and core temperature T_{core}, which can be linked to thermal sensation model to project the human's thermal sensation and thermal comfort at local and overall body level. In our approach, we trained a Machine Learning-based human thermo-physiology model ML-HTPM, based on the data from the physical HTPM, to predict the local skin, mean skin, and core temperature using features that can be extracted from RGB images. Fig. 4 presents the steps used to develop the ML-HTPM which was based on the dataset generated from a physical HTPM. The input of environmental parameters was replaced by the head and hand skin temperature as features to calibrate the ML-HTPM.

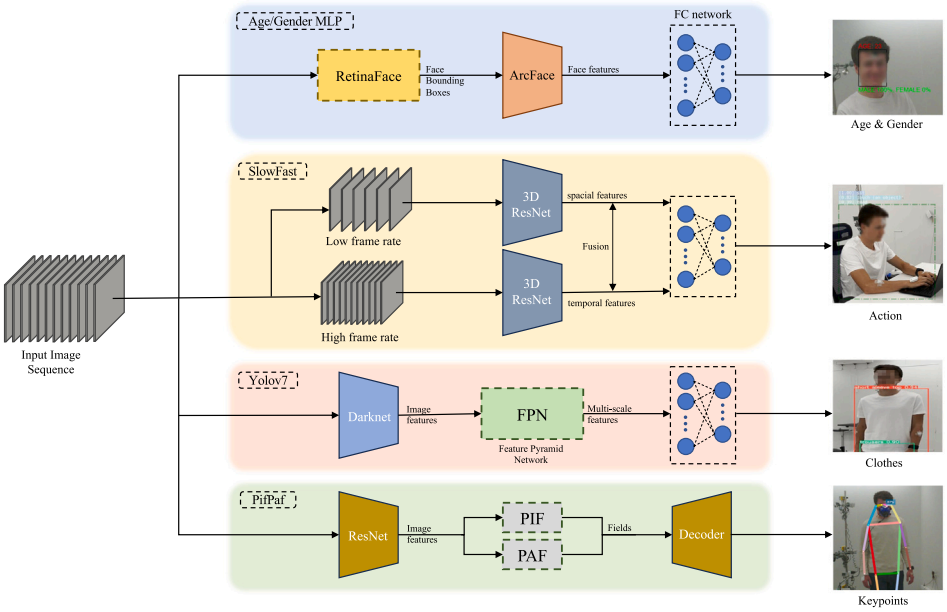

Fig. 2. Overview of the computer vision models for personal features extraction.

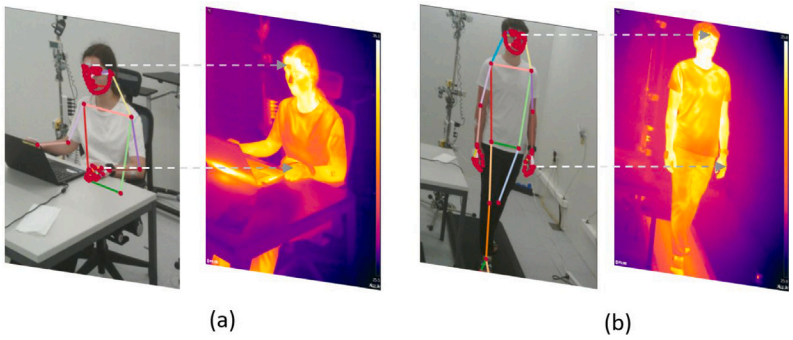

Fig. 3. Body parts detection in the RGB image using OpenPifPaf and the link to the body temperatures in the IR image.

2.2.1. Data generation

The HTPM JOS3 was used to generate two databases that served for training and testing the ML-HTPM. In order to generate a reliable and explicit training dataset, the following steps were undertaken:

1. 20 human subjects (10 males and 10 females) with actual measured anthropological data such as height, weight, and body fat percentage from studies [31,32,70] were used for the simulations. The mean and standard deviation of anthropological data of the considered population are provided in Table 1.
2. For each subject, a sequential set of simulations was conducted by exposing the human model to step-changing operative temperature while looping over the other personal and environmental parameters. The time used between changes in temperature steps was 30 min, allowing for 60 min over the thermal neutral temperature of 24 °C. Fig. 5 shows the step change operative temperature profile used in the simulations, also Table 2 presents the values of personal and environmental factors that were considered for the dataset generation.
3. All data were combined and arranged by subjects forming one dataset referred to as *training dataset*. The training dataset including data from all simulations reached the size of 5.8M where the line of data represents consecutive minutely data.

For testing purposes, two kinds of datasets were generated using a physical JOS3 for 4 familiar (already used in the training data) and 4 unfamiliar (i.e., new) people with anthropological data different than what was used for training. The first dataset named *"dynamic testing"* followed the same environmental profile as during the training phase. The second dataset named *"varying activity testing"* aimed to challenge the model. In a new testing dataset, activities were frequently changing, and corresponding metabolic rate and air speed around the body parts were dynamically varying as presented in Fig. 6. The profile was chosen based on the experiment presented in [31]. Simulations for *"varying activity testing"* were conducted over 4 uniform and steady operative temperatures of 22, 24, 26, and 28 °C over 215 min each.

Table 1
Anthropological data of the population considered in the training dataset.

	Height [m]	Weight [kg]	Fat [%]	Age [Y]
10 Males	1.74 ± 0.10	71.08 ± 8.4	19.39 ± 5.4	31.2 ± 6
10 Females	1.65 ± 0.05	61.85 ± 7.3	26.92 ± 7.0	29.1 ± 12

Table 2
Environmental and personal factors used to generate the training dataset.

Factors	Metabolic rate [met]	Clothing [clo]	Relative humidity [%]	Air speed [m/s]
Values	1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6	0.4, 0.6, 0.7, 0.8	40, 60	0.1, 0.5

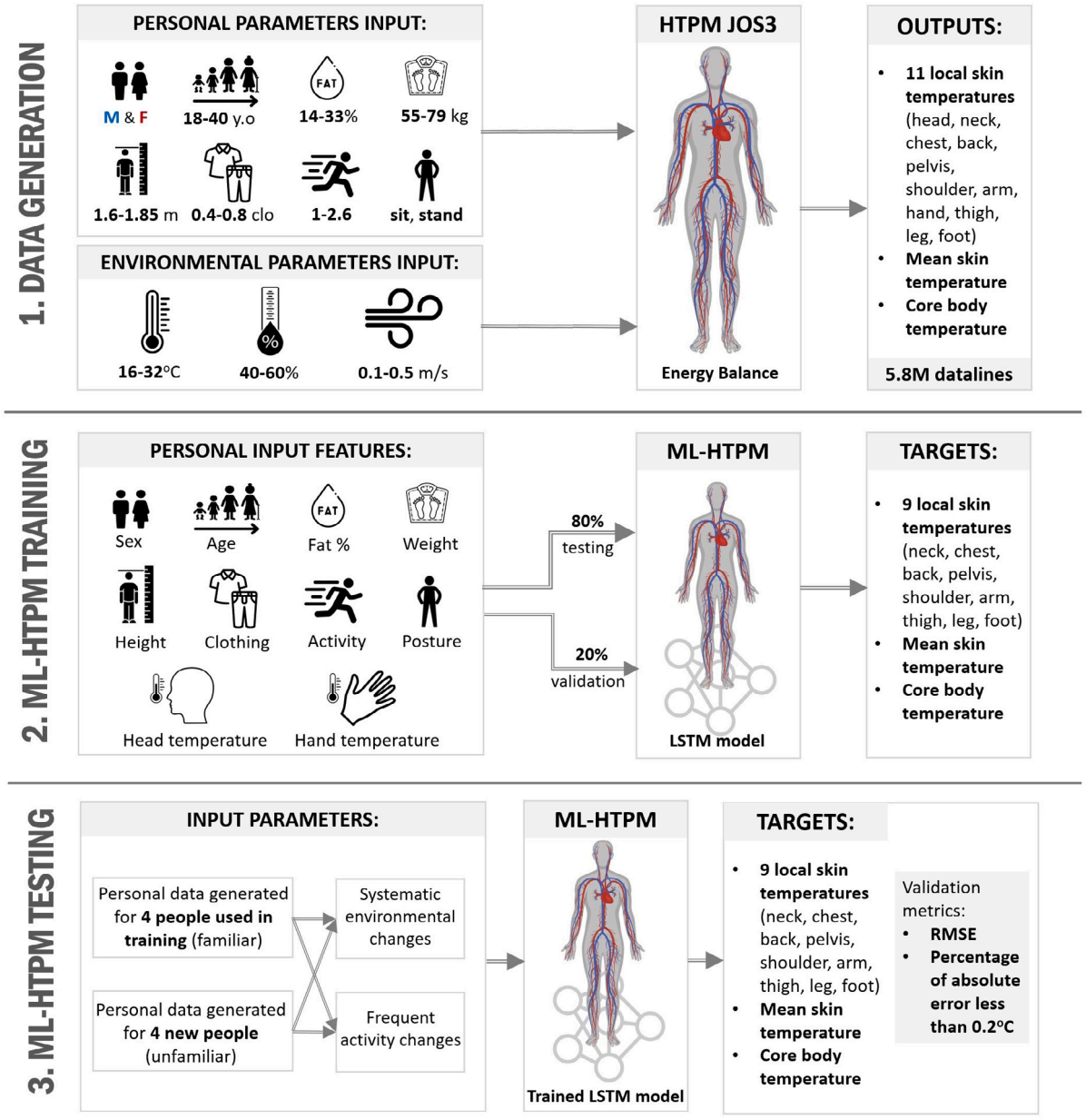

Fig. 4. Development steps of Machine Learning-based human thermo-physiology model (ML-HTPM).

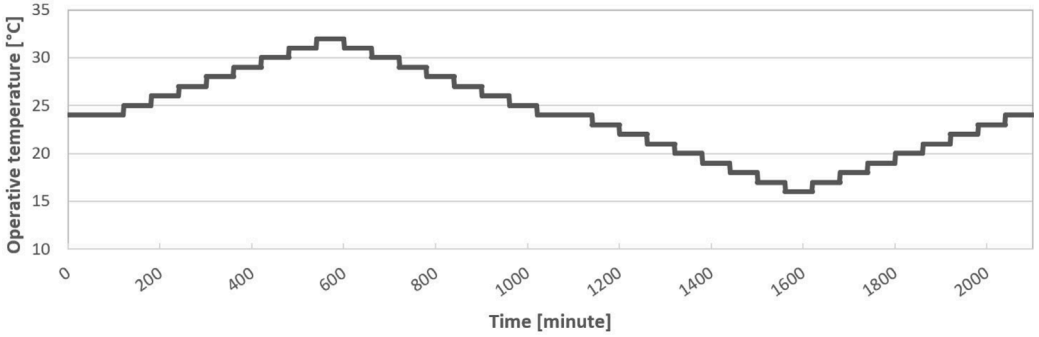

Fig. 5. Operative temperature profile in simulations to generate a training dataset.

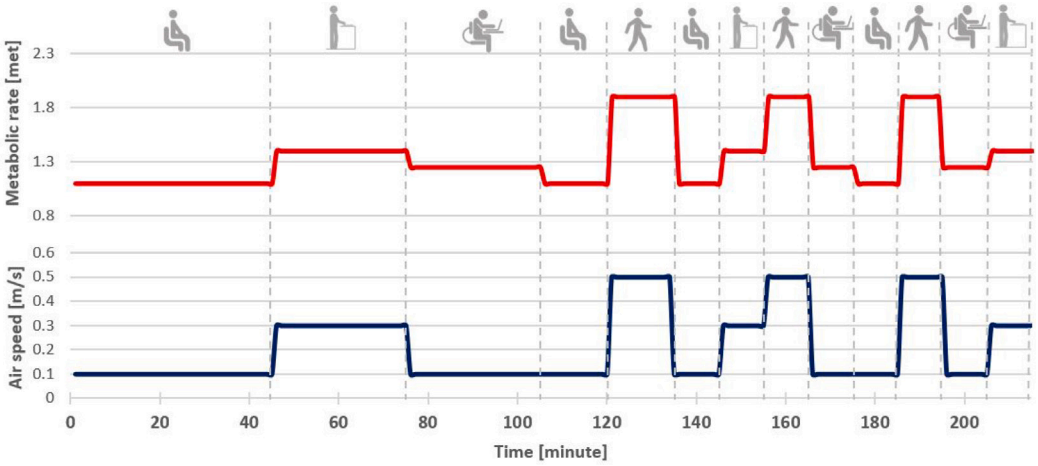

Fig. 6. Metabolic rate and air speed variation for generation of testing dataset at each fixed operative temperature case.

2.2.2. Feature selection for the development of the ML-HTPM

The generated training dataset was re-processed, keeping the data arranged in a time flow. The features and targets were identified and arranged to meet the objective of the ML-HTPM. The environmental parameters were not used in ML-HTPM however, the model relied on the use of two local skin temperatures of the head and the hand. Two variations of the ML-HTPM model were developed. One model was based on 10 features selected for the training:

- *physiological features* (two local skin temperatures T_{head} and T_{hand})
- *personal features* (clothing insulation *Clo*, metabolic rate *Met*, posture *P*)
- *individual body composition* (age *A*, sex *S*, height *H*, weight *W*, fat percentage *F*).

Due to the difficulty of estimating the height, weight, and fat percentage of a human from an RGB image without the use of a depth camera, the second variation of the ML-HTPM excluded these features and was only based on the remaining 7 features. The aim of developing two variations of models was to understand the influence of body composition-related parameters on prediction accuracy. The features selected to predict the thermal state of individuals can be summarized as follows:

$$F_7 = [T_{head}, T_{hand}, Met, Clo, P, A, S] \quad (1)$$

$$F_{10} = [T_{head}, T_{hand}, Met, Clo, P, A, S, H, W, F] \quad (2)$$

2.2.3. ML-HTPM modeling architecture using LSTM networks

The ML-HTPM should consider the transient nature of the human thermo-physiology model and the fact that metabolic rate, clothing, and environmental exposure are dynamically changing. One of the key challenges in handling sequential data is retaining information from earlier time steps and effectively utilizing it when processing subsequent steps. Therefore, we developed a time series regression prediction model based on Long Short-Term Memory (LSTM) to estimate the temperatures of different human body parts while selected features are used as inputs. LSTM is a type of recurrent neural network (RNN) that is well-suited for modeling sequential data due to its ability to capture long-term dependencies [71]. In our model architecture, we design a four-layer LSTM module, each layer containing 16 LSTM units, which takes the input features across a time window of 10 s and learns their temporal relationships. The LSTM is followed by a fully connected neural network (Fig. 7) with 256 hidden neurons, employing ReLU activation functions for non-linearity. The final output layer predicts the temperatures of the 11 body parts (neck, chest, back, pelvis, shoulders, arm, thigh, leg, foot, and core body). We train the model for 300 epochs using a batch size of 128, the learning rate of 10^{-3}, and Adam optimizer. As the LSTM algorithm is trained based on the variations in the data over a certain number of previous timesteps, intervals of 10 and 20 min were evaluated. By leveraging the power of LSTM and a deep full connected network, our model aims to provide accurate predictions of body temperatures based on the given input factors.

3. Results

3.1. Performance of features extraction models from RGB images

The computer vision models that were chosen had acceptable performance when tested with their respective datasets. The age and gender detection model achieved a classification accuracy of 90.66% on the test set of Audience [51]. The action classification model achieved a top-1 and top-5 classification accuracies of 79.8% and 93.9%, respectively [63]. Similarly, the clothing detection also achieved good accuracy when using YOLOv7, the model achieved an Average Precision (AP) score of 82.4%. The OpenPifPaf had an AP score of 71.9% on COCO 2017 TEST-DEV dataset [69].

The feature extraction techniques were applied to images of subjects from our own experiments presented in [31]. Fig. 8 illustrates the output after applying age and gender detection, the model detects the gender accurately and age with an acceptable error within a range of ±2 years in this particular case. Fig. 8 shows an example of a female (27 y.o.) and a male (22 y.o.) where the model was applied to side and front face profile images for the same people. An example of the prediction of people's activity is shown in Fig. 9. Image (a) shows a sitting person performing office work (site, touch an object, read) with a metabolic rate corresponding to 1.2 met. A walking person (walk) with a corresponding metabolic rate of 1.7 met is shown in the image (b), and image (c) shows a standing person (standing, carry/hold) with a corresponding metabolic rate of 1.4 met.

Example outputs of the clothing detection model are shown in Fig. 10. Two images show a standing person, one has a full-body appearance while the second shows only the upper half of the body. Since the model can only predict what it can see, it showed only what the person is wearing at the top in the second case. However, for the remaining seated cases, the model was able to predict the clothing of both upper and lower body parts even though the person was partially obstructed by the desk. The model detected the short-sleeve and long-sleeved shirts and the trouser correctly. Apart from recognizing the clothing contours, the model does not provide information about the

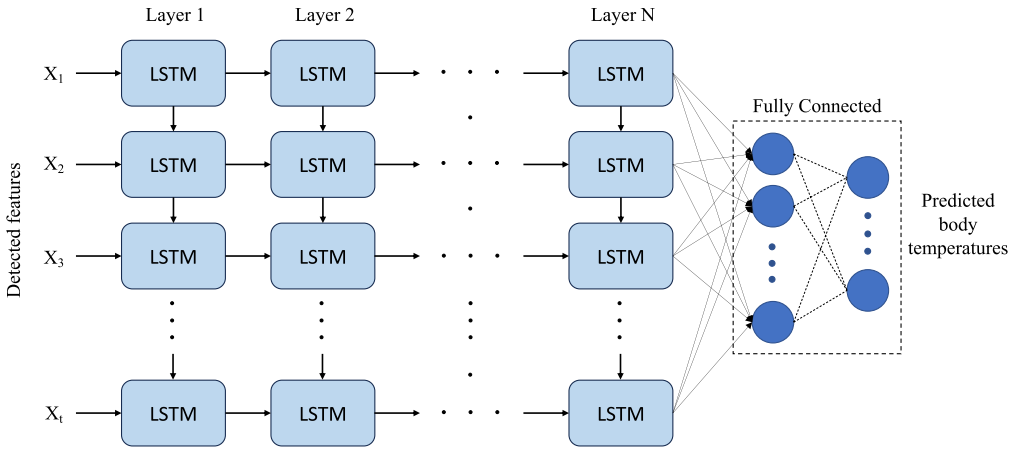

Fig. 7. LSTM regression model architecture.

Fig. 8. Example illustrations of age and gender detection (images taken from controlled experiments [31]).

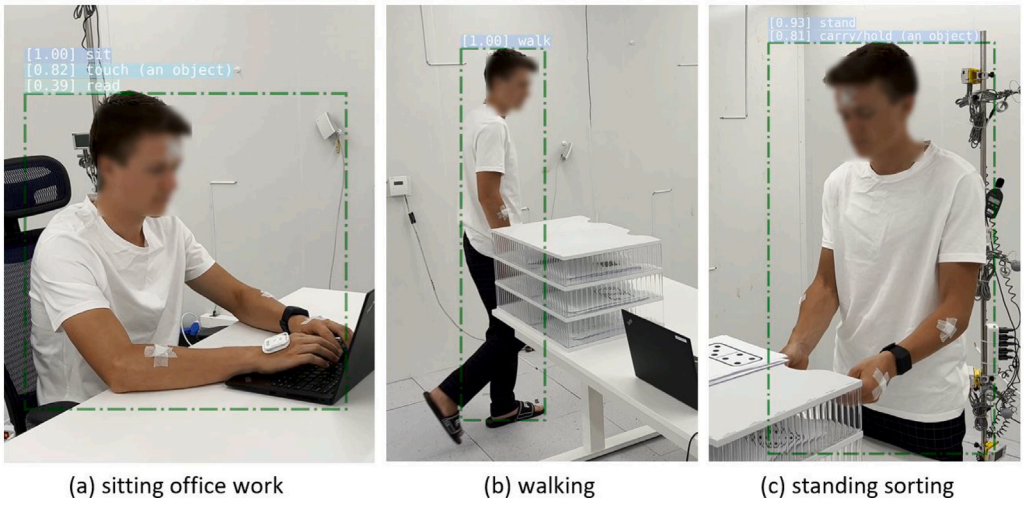

Fig. 9. Examples of actions recognition (images taken from controlled experiments [31]).

material of clothing, thickness, or air gap, which brings some challenges in detecting the actual insulation value. For example, as the model predicts both the jeans and the cotton trousers as trousers, both would refer to an insulation value of 0.24 clo (Fig. C.19).

3.2. ML-HTPM results and performance

First of all, the general performance of the ML-HTPM model using 7 and 10 features and also LSTM sequences of 10 min and 20 min was evaluated. The purpose was to see the influence of the number of features and the effect of the sequence time on the performance of the model. The results on selected skin temperatures and core body temperature prediction are presented in Fig. 11 for *dynamic testing* in (a–b) and for *varying activity testing* in (b–c). In addition, two datasets with 4 people were used: (i) 4 people from the training dataset that were *familiar* to the model (a, c), and (ii) 4 new people *unfamiliar* for the model (b, d). Due to the uniform environment consideration in the data generation, HTPM predicts symmetrical temperature for left and right body parts; thus, we considered the left and right extremities as one body part.

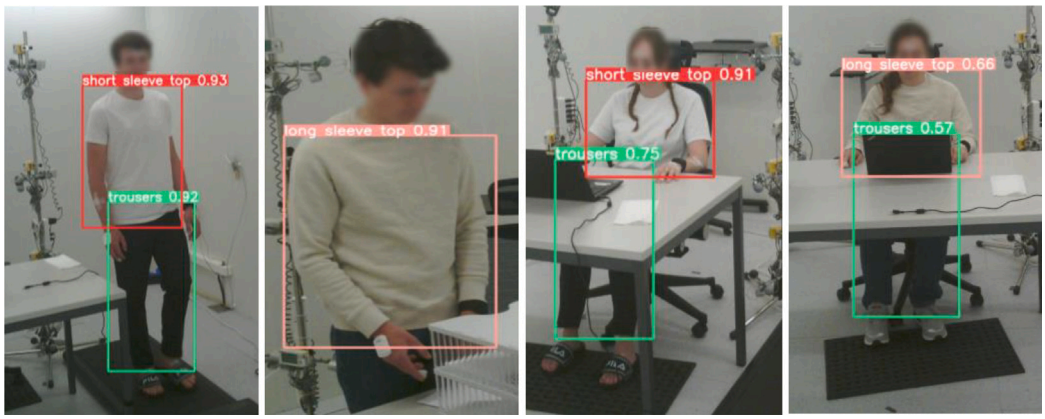

Fig. 10. Examples of clothing classification with the indication of the prediction confidence (images from study [31]).

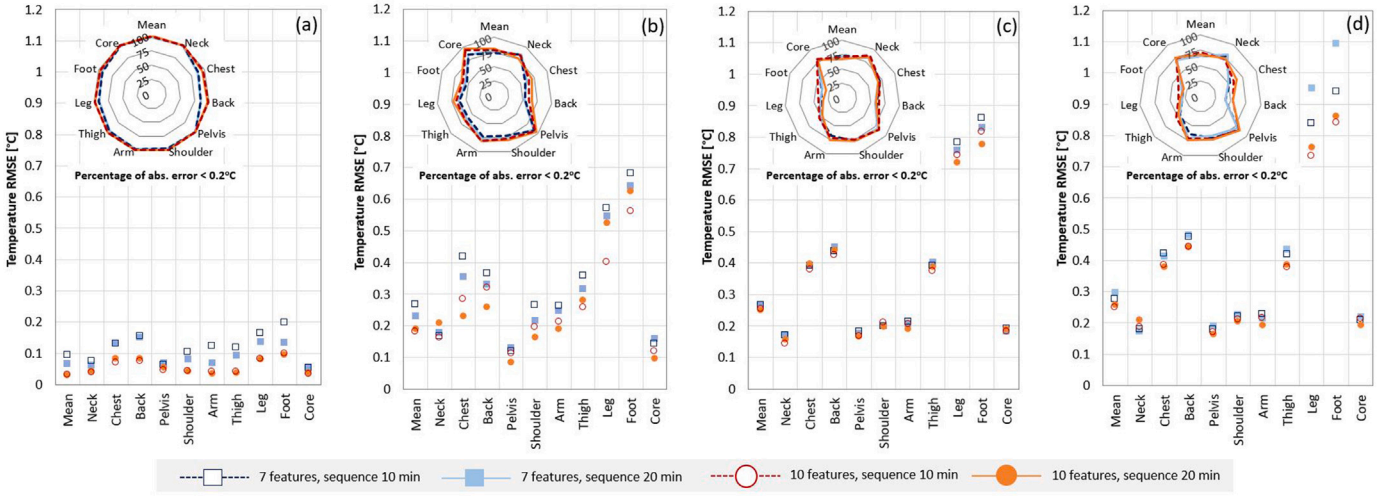

Fig. 11. Performance of the ML-HTPM in terms of RMSE of temperature and percentage of absolute error less than 0.2 °C for the four different cases: (a) *dynamic testing* of 4 familiar people, (b) *dynamic testing* of 4 unfamiliar people, (c) *varying activity testing* of 4 familiar people, (d) *varying activity testing* of 4 unfamiliar people.

The results show that increasing the number of features by adding height, weight, and fat percentage improves the prediction accuracy, especially for data generated from new people. The percentage of absolute error (PAE) < 0.2 °C increased for both data from *familiar* people and *unfamiliar* new people (Fig. 11a vs. b) which means that the detailed body composition might need to be considered to have a better prediction of the model applied on a larger population. The results showed an increase in PAE of around 7% and around 0.1 °C improvement in RMSE in some body parts. The effect of the number of previous timestamps on prediction can be clearly seen in Fig. 11(d). The model with the smaller sequence of 10 min tends to have a better performance when applied to a varying activity testing dataset. Based on the results, the performance increased by 8% in some body parts when decreasing the sequence from 20 min to 10 min on the varying activity testing dataset; however, it did not show an effect and sometimes negative performance when applied to the dynamic testing dataset. Both leg and foot skin temperature showed the highest error with an RMSE reaching 1 °C in the worst case of the new people and varying activity testing dataset. To better understand the increased error in extremities, the distribution of temperatures at different body parts in the training dataset from a male and a female is presented in Fig. 12 presents. The violin plots show that extremities (e.g. hand, leg, and feet) had a wider range of temperatures, and core body parts had a smaller range of temperatures. The extra data concentrated at low temperatures, for example in the feet, is due to different clothing insulation which is not shown in nude body parts such as the forehead and hand.

To show how well the ML-HTPM predicts temperatures in a dynamic situation, Fig. 13 presents the dynamic prediction of mean, core, foot, and chest skin temperatures from four different environmental temperatures (22–28 °C) while a person was frequently changing activity according to illustrations in Fig. 6. The results of ML-HTPM shown in the figure are the prediction of the 7 input features and 10 min sequence for a new person (unfamiliar to the model). The results for the complex scenario show relatively acceptable performance, especially for the mean skin and core temperatures with a small error. The discrepancy occurs mostly during standing and walking activity, which is accompanied by sudden elevated metabolic rate and increased air speed. The dataset generated for training did not include data with a sudden change in metabolic rate, instead, the change of metabolic rate or air speed was set once before each simulation covering all the combinations of metabolic rate, air speed, relative humidity, or clothing insulation.

4. Discussion

The ML-HTPM developed in this paper shows the potential of using a machine learning algorithm to predict the physiological adaptation to

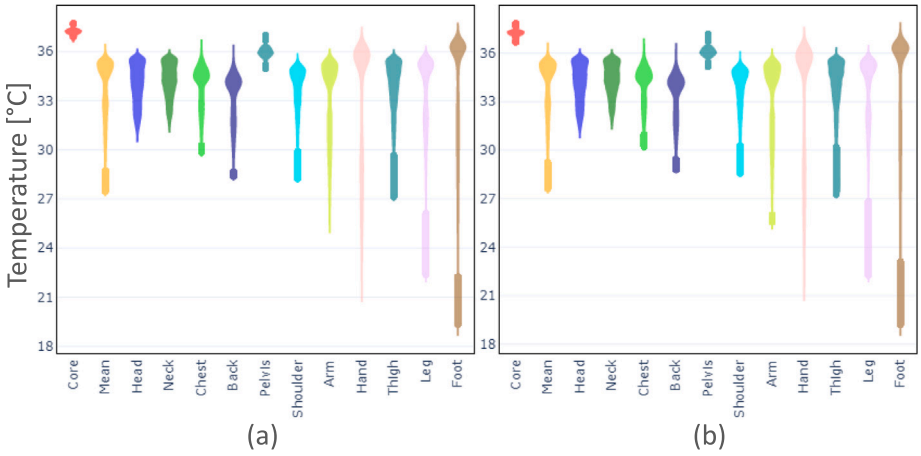

Fig. 12. Example distributions of local skin, mean, and core temperatures for a male and a female subject.

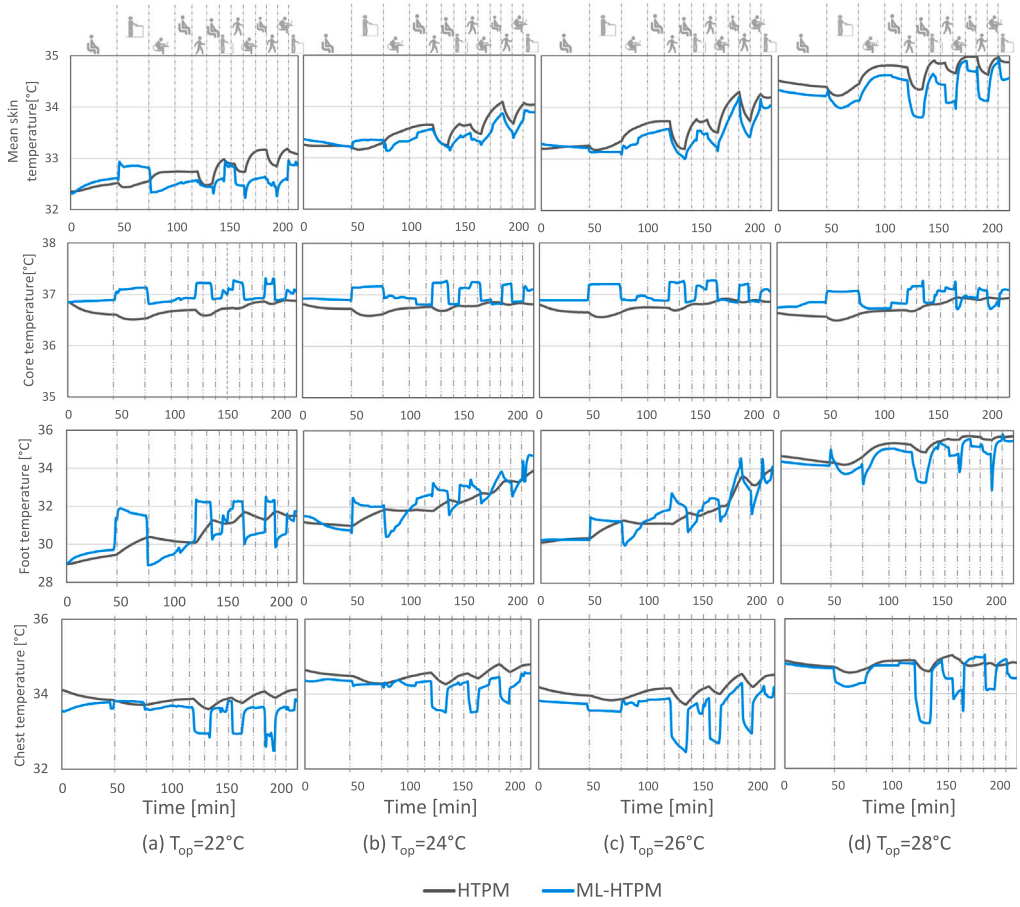

Fig. 13. Dynamic performance of ML-HTPM with 7 input features and 10 min sequence in four different environmental conditions: (a) at 22 °C, (b) at 24 °C, (c) at 26 °C, and (d) at 28 °C for mean skin, core, foot and chest temperatures.

the environmental changes of an individual by relying on a few inputs, thus, this tool could serve the purpose of sensing an individual's thermal states with minimum intrusive measurements. By further analyzing the predicted thermo-physiological parameters of local skin, mean skin, and core temperatures, it would be possible to evaluate the person's thermal sensation. In real life, human tends to be exposed to a dynamic environment where people frequently change activity and clothing. Therefore, this paper considered the influence of a dynamic environment in the development of ML-HTPM by considering the LSTM model for training. Based on the results, the model with the smaller number of previous timestamps tends to have a better performance when applied to a frequently changing environment. The results showed that 10 min of historical data can be sufficient to improve predictability and to account for the dynamic variation in skin temperatures.

In ML-HTPM, the hand and head skin temperatures seem to have a strong correlation with adjacent body parts, as the error from the adjacent body parts is minimal. However, the leg and foot temperatures exhibit the highest error, which can be attributed to the fact that these body parts are the furthest from the head and hand. As the distance

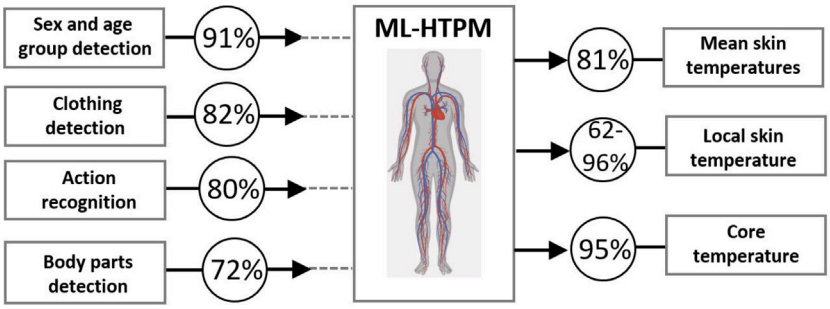

Fig. 14. Flowchart of the proposed method and accuracy of different components.

between body parts increases, the strength of this relationship diminishes slightly. Furthermore, increasing the number of features resulted in a minor improvement in Root Mean Square Error (RMSE) for certain body parts. This implies that models should be trained on datasets generated from a more diverse population to account for variations in physiological responses related to individual body composition.

Moreover, the model was not trained on data where the air speed or metabolic was suddenly changing, instead, the training dataset was based on a fixed metabolic rate, air velocity, etc., while only the operative temperature was dynamically changing. The influence of sudden change in air speed and metabolic rate was shown in the dynamic results in Fig. 13; therefore, improving the predictability of the model can be achieved by adding more simulation cases to the training dataset where a person experiences a sudden change in air speed and metabolic rate. Based on the results from ML-HTPM, we saw that including information about an individual's body composition such as height, weight, and fat percentage increases the accuracy (PAE) by around 10%. Extracting those features from images requires the use of a depth camera or other advanced techniques which was outside the scope of this work, but should be considered in future research.

The methods presented in this paper require the use of a multimodal sensing and machine learning. Fig. 14 summarizes the prediction accuracies of all computer vision models reaching the ML-HTPM and the accuracy of the ML-HTPM model outputs if the inputs are correct. Results from ML-HTPM showed a PAE of 81% for mean skin temperature, 95% for core temperature, and a range of (62%–96%) for the different local skin temperatures. Those percentages of performance are for the ML-HTPM receiving accurate inputs, the precision is expected to drop if the inputs are directly taken from the computer-vision models. Although the computer vision algorithms adopted to extract personal thermal comfort-defining features from images showed good accuracy, they cumulatively might reduce the accuracy of the ML-HTPM model. Therefore, the accuracy of the feature extraction should be maximized to maximize the accuracy of the ML-HTPM prediction.

Specific issues related to computer vision models are the following. Prediction of clothing insulation as well as the metabolic rate estimation are based on the values prescribed in standards. In reality, these values, particularly metabolic rate, might be subject to inter-individual differences [32]. The clothing detection model can provide a good description of the clothing items; however, it misses detailed features of clothing that can help estimate the thermal resistance values more accurately. The garment's materials, thickness, color, and air gaps have a significant influence on defining the clothing's insulation [72,73], which cannot be extracted from an image yet. The clothing detection model has an accuracy of 82% which describes the accuracy of the model in classifying the type of clothing based on the shape. Some body parts might not be in the field of view of the camera due to obstruction; however, this can be overcome by using face recognition and tracking and trying to predict clothing whenever the person has a full view. The action recognition requires a sequence of images to predict, and the model showed an accuracy of 80% (top-1). The model gives information that describes the person's activity, which is sufficient only to estimate a metabolic rate value derived from the standard descriptions. Both age and sex prediction model has a 91% of accuracy and from the tests conducted on images from our experiments, the model showed good agreement with real values. It was predicting the sex correctly, however, the absolute age showed some variation for the same person from different images and profile perspectives, but the age group prediction was correct. OpenPifPaf has adequate accuracy in locating the person's body parts with 72% accuracy; however, it requires both hand and face to determine keypoints and draw a human body skeleton. Thus, in cases when a person's face is obstructed, it would be challenging to detect the body parts and further link them to the temperatures from the IR camera. In many practical cases it might be difficult to have both a face and a hand of the same person simultaneously in the field of view of cameras. This could be technically solved by using a rotating motor and slides for the camera or intermittent monitoring. Otherwise, further analysis on what extent the modeling accuracy could drop if only one body part is captured should be performed.

Finally, the ML-HTPM is trained on data generated from a physical HTPM JOS3, a model based on energy balance equations that consider physiological phenomena. Factors such as ethnicity, thermal acclimatization, and circadian rhythm were not implemented in those models yet. Also, JOS3 has some issues with personalized thermal parameters prediction, particularly in extremities, when compared with actual data. Based on our previous study [31], the RMSE can reach 3 °C in the foot, 0.9 °C in mean skin temperature, and 0.3 °C in core temperature. The ML-HTPM was not yet evaluated with real data; however, it showed promising results when compared to HTPM results with an RMSE less than 0.5 °C in most of the body parts.

As the suggested framework involves multiple computer vision models for data collection, some privacy challenges arise. As a solution, the models should be performed in real-time, with no storage of images in the local or cloud storage; thus, the privacy of people during the data acquisition will be assured at the monitoring time. Only the numerical data corresponding to the ID of the person in the field of view can be translated to the ML-HTPM model that will ultimately output their thermal state parameters.

5. Conclusion

Contactless monitoring of the comfort of people can be used for surveying if the building meets comfort criteria and for better climate control of buildings. The camera-based solution can provide input in terms of the actual thermal state of people to the controls of a climatic system, both centralized and personalized. The potential of monitoring at the individual level will advance the implementation of Personalized Environmental Control Systems (PECS), the new generation of technologies designed to condition the micro-environment around humans, thus avoiding energy waste to condition the spaces that are not occupied.

With the improvement in the field of data science and machine learning, it becomes feasible to predict an individual's thermal state

through data-driven approaches. Accordingly, this paper showed that machine learning can emulate individuals' thermo-physiological state (local and mean skin temperature, core temperature) by considering a few physiological inputs and some personal factors. The multi-modal method presented in this paper also showed that we can extract all needed personal thermal comfort-defining features from RGB and IR images. Based on the results of the developed machine-learning-based human thermo-physiology model (ML-HTPM), the model was able to predict the temperature distribution over the body with an RMSE varying between 0.2 and 1 °C in the worst case. The highest RMSE was shown at the leg and the foot, the body parts that are away from the hand and head, which were considered as physiological inputs. The models used to extract features from RGB images exhibited high accuracy in predicting personal features, with a clothing detection accuracy of 82%, an action recognition accuracy of 80%, and an age and sex recognition accuracy of 91%. Moreover, the method uses IR images with the help of a machine learning pose estimator OpenPifPaf to extract skin temperatures of the hand and head non-intrusively in real-time.

Considering the dynamic variability of environmental parameters and people's diversity in the training dataset can improve the prediction accuracy of the ML-HTPM which should be considered in the further development of the proposed framework. Generally, an unlimited combination of environmental and personal factors can be considered to generate the data and generalize the model. In the development of the framework, we restricted ourselves to personal characteristics of individuals that were actually measured and to environmental parameters typical for offices. The further note that the model was trained on a dataset of uniform temperature exposure over all body parts, thus, it needs to be re-trained for cases when an asymmetric environment or localized heating or cooling is introduced using HTPM outputs for such conditions. It is important to make sure that the source HTPM is validated for non-uniform environments. More investigation is needed to study the applicability of the ML-HTPM in such a non-uniform environment. In general, the ML-HTPM proved its functionality in projecting an individual's thermal state; thus, training a model on measured data can present a more realistic prediction, however, this requires an extensive dataset on human thermal responses.

Overall, the presented framework shows that a human can be used as a sensor strengthening the need to move toward an occupant-centric approach, which can be further used to control the micro-environment of individuals. Developing a human-centered indoor climate based on continuous monitoring of occupants' comfort can improve the quality of living, work performance, and overall satisfaction with the built environment.

CRediT authorship contribution statement

Mohamad Rida: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Mohamed Abdelfattah:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Alexandre Alahi:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition. **Dolaana Khovalyg:** Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dolaana Khovalyg reports financial support was provided by Federal Polytechnic School of Lausanne.

Data availability

The authors do not have permission to share data

Acknowledgments

This work was supported by the ENAC Interdisciplinary Cluster Grant 2021, and it contributes to IEA EBC Annex 87 "Energy and Indoor Environmental Quality Performance of Personalised Environmental Control Systems". We thank all the participants who volunteered for the experiments and carefully followed the procedures.

Appendix A. Human thermo-physiology model JOS3

Model JOS3 segments the human body into 17 parts [23]. Each body part comprises multiple concentric layers (core, muscle, fat, skin, artery, and vein), forming 83 nodes in total. The model is based on the energy balance between the environment and the human body; thus, it includes heat transfer between the skin layer and the environment through *3 modes (convection, radiation, evaporation)* in addition to *conduction* between the different layers and *convection* due to the blood circulation between layers and body parts [23], as illustrated in Fig. A.15. JOS3 included the vein and artery nodes. The model incorporates a sophisticated interconnected blood flow system, comprising nodes representing arteries and veins within each body part. Additionally, it includes a superficial vein in the extremity. The distribution of arteries and veins across the various body parts is determined based on the formulation described in the work of Smith [74]. JOS3 also included the modeling of arteriovenous anastomoses (AVA) blood circulation phenomena in hands and feet to improve the overall model

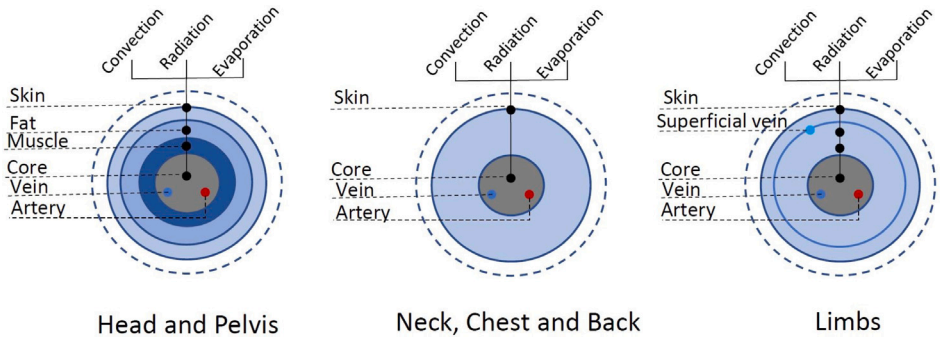

Fig. A.15. Heat transfer and nodes representation of JOS3.

predictability. JOS3 considers different thermoregulatory mechanisms (vasodilation, vasoconstriction, sweating, and shivering) by using feedback temperature sensed by thermoreceptors located throughout the body parts. Similar to other HTPMs, JOS3 requires local environmental parameters, personal parameters, and individual body characteristics as input.

Appendix B. Evaluation of temperature measurements using an IR camera

To evaluate the usefulness of using the measurements from an IR camera, we have conducted a series of controlled experiments and compared IR measurements with contact skin measurements. The experiments were conducted in a climatic chamber monitoring the skin temperatures of both hands and forehead for six subjects (3 males M1–M3 and 3 females F1–F3) using in-house calibrated iButton skin temperature sensors (±0.2 °C) and FLIR A700 infrared camera (±2 °C). All experiments started at around thermal neutrality after that, each subject was exposed to four different environmental conditions [22–24–26–28 °C] over 3 h and 35 min on separate days. During the experiment, the subjects conducted a sequence of standardized office activities and wore standard summer clothing (0.35clo) at 26 and 28 °C and winter clothing (0.65clo) at 22 and 24 °C. The activities conducted during the experiment have the same flow as in Fig. 6. The thermal camera FLIR A700 used in the experiment is one of the advanced monitoring tools developed by FLIR. The lens used was with a focal length of 10 mm (42 °C) to have a larger field of view, capturing all the activity of the subject. The camera produces images with 640 × 480 pixels, and with the measurement range, the manufacturer stated an accuracy of ±2 °C. Fig. B.16 shows the approach used to evaluate the IR measurements by identifying the pixels on the IR image close to the contact iButton sensor.

The results presented in Figs. B.17 and B.18 show the data from the three males and three females, respectively. Each figure presents four different experiments showing the IR measurements and data from iButton for the forehead, the hand, and its difference. It shows that in some cases IR can predict the skin temperature very well, however in some cases the error reached ±2 °C. The difference between hand and head skin temperature might provide less error, since the IR measurement error might be applicable to all pixels in one frame and FLIR does an auto adjustment and calibration regularly during measurement.

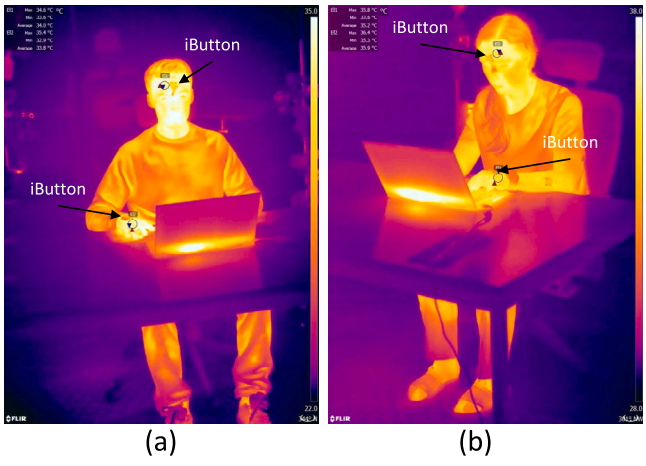

Fig. B.16. IR images from the experiments showing examples of the pixels considered for temperature readings and the comparison with iButton sensors.

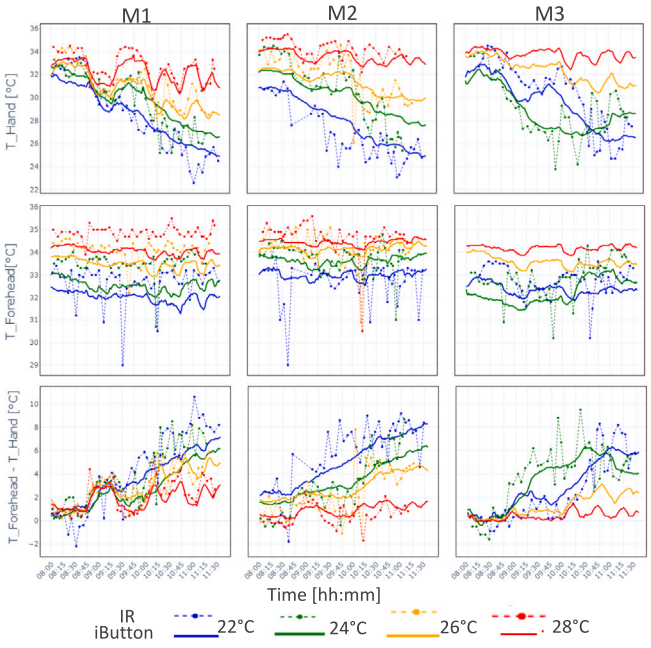

Fig. B.17. A comparison between the IR measurements and the on skin iButton sensor temperature measurements for the head and hand from three males at four different experiments.

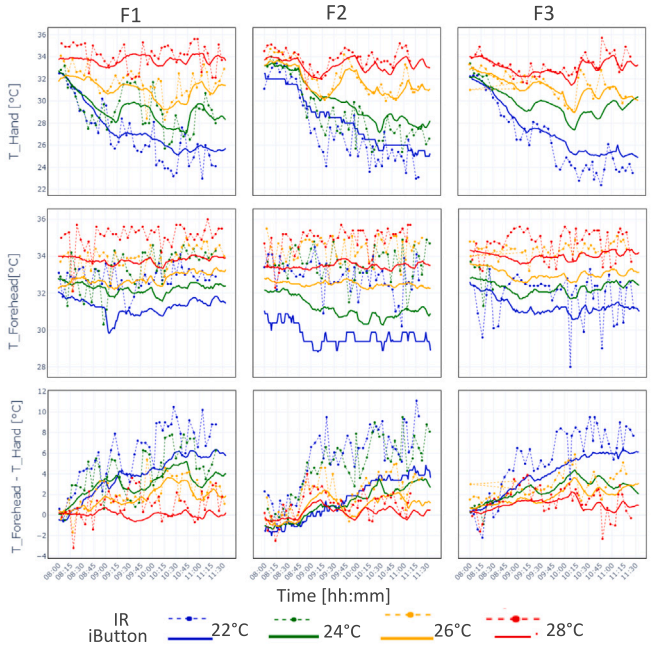

Fig. B.18. A comparison between the IR measurements and the on skin iButton sensor temperature measurements for the head and hand from three females at four different experiments.

Appendix C. Description of computer vision models for personal features extraction

In this section, a detailed overview of the computer vision models used for the extraction of personal features from RGB images is presented. It includes age and gender detection using a state-of-the-art model trained on the IMDB-WIKI and Adience datasets. Action

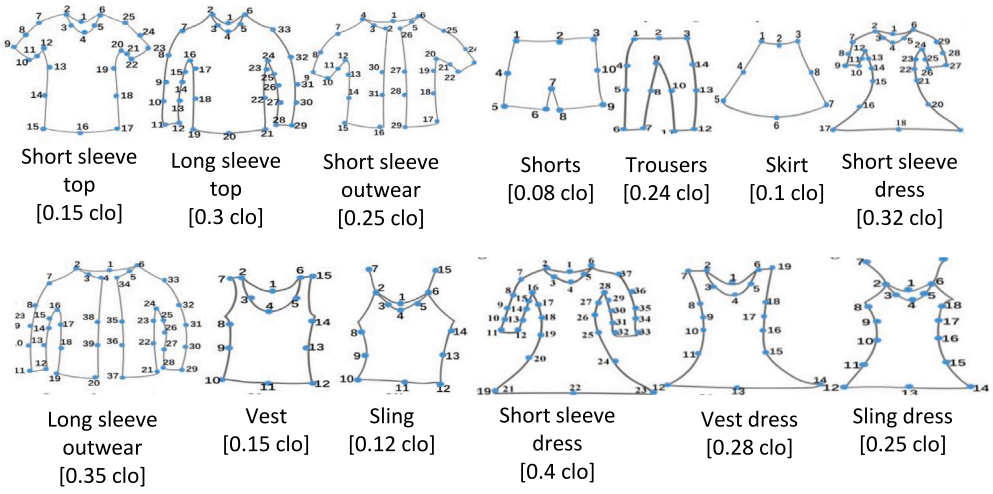

Fig. C.19. Clothing classification in DeepFashion2 and the matched values of clothing insulation (in clo) from ISO7730 [68].

classification utilizes the SlowFast network trained on the Kinetics-700 dataset, achieving high accuracy in recognizing various activities. Clothing detection relies on the YOLOv7 model trained on the DeepFashion2 dataset, showcasing improved performance with dynamic anchor assignment and feature pyramid integration. Finally, posture and body parts detection employ the OpenPifPaf library, which accurately estimates human pose by locating key points.

C.1. Age and gender detection

We implemented the model proposed by [51] due to its state-of-the-art performance in age and gender classification. From input RGB images, the model predicts both the age and gender of every person in the image through a three-stage process. First, the image is fed to a pre-trained RetinaFace [65] model, the current best model for face detection, to get the bounding box of the face of each person. The bounding boxes are used to crop the faces out of the image, hence removing background noise and reducing data size. Secondly, RGB faces are passed to a powerful face-embedding extractor, ArcFace [66], to extract the features and further reduce the data size. Finally, a Multilayer Perceptron (MLP) is applied to the embeddings to ultimately predict age and gender. The MLP consists of a fully connected layer, followed by ReLU, batch normalization, and dropout layers (cascaded five times). The model has been pre-trained on the popular IMDB-WIKI dataset [75], which contains 523 K images of celebrities annotated with gender (two classes: male, female) and fine-grained 101 age categories (*i.e.* from 0 to 100). It was then fine-tuned on the Adience dataset [76], which encompasses 19 K images annotated with age and gender. It achieves a classification accuracy of 90.66% on the test set of Audience.

C.2. Action classification

The SlowFast [63] network is used to extract information about human activities from input video streams. The network is composed of two parallel pathways: a slow pathway, operating at a low frame rate, to capture special semantics, and a fast one, operating at a high frame rate, to capture temporal information. We use the default instantiation proposed by the authors, with the 3D-ResNet as the backbone network for both pathways. The two pathways are fused by lateral connections into a SlowFast network. The model has been trained on the Kinetics-700 [77] dataset, which contains 545 K training videos and 30k validation videos. Actions are classified into 700 categories which include walking, sitting, running, carrying an object, throwing, playing, etc.

C.3. Clothing detection

To automatically detect the types of clothes of each person in the image, we run YOLOv7 [56], the state-of-the-art object detection model, on input RGB images. The model ultimately predicts both the classes and bounding boxes of different items of clothes in each image. The YOLOv7 network is composed of a series of convolutional layers that downsample the input image, followed by several detection layers that predict the bounding boxes and associated class probabilities for a fixed number of predefined anchor boxes. Each anchor box is defined as a prior box that is centered on a specific location in the image and has a certain width and height.

During training, YOLOv7 optimizes the network parameters by minimizing a loss function that penalizes errors in the predicted bounding box coordinates and class probabilities. The loss function is composed of several components, including the mean squared error (MSE) between the predicted and true bounding box coordinates, the cross-entropy loss between the predicted and true class probabilities, and a regularization term that encourages the model to learn sparse representations.

YOLOv7 introduces several improvements over previous YOLO models, such as the use of a Swish activation function, which has been shown to improve the accuracy of deep neural networks, and the integration of a feature pyramid network (FPN) [78] to capture multi-scale features across different layers of the network. Additionally, YOLOv7 utilizes a dynamic anchor assignment strategy to adapt the anchor boxes to the object sizes and aspect ratios in the training dataset.

We trained the network on DeepFashion2 [67], the largest publicly available dataset for clothes detection. DeepFashion2 contains 492 K images, 873 K clothing items, 801 K bounding boxes, and 13 clothes categories. Fig. C.19 shows examples of the clothing categories that can be identified by the model and the estimated clothing insulation values taken from ISO 7730 [68].

C.4. Posture and body parts detection

For the identification of specific body parts and posture from an image, we considered the machine learning library OpenPifPaf [69]. The library incorporates a recent model that offers detailed keypoint annotations for the face, hands, and feet, comprising a total of 133 keypoints [79]. This model enables accurate human pose estimation with fine-grained keypoint information. The architecture of OpenPifPaf is based on a generic neural network that detects and constructs a spatio-temporal pose. This pose is represented by a connected graph, where each node corresponds to a person's body joint across multiple frames.

Appendix D. Metabolic rate values adopted from standards

Metabolic rate values used in this work were extracted from ASHRAE 55 standard (Table D.3).

Table D.3
Examples of office work metabolic rate values from ASHRAE 55 [80].

Activity	Values in Met units
Resting	
Sleeping	0.7
Reclining	0.8
Seated, quiet	1.0
Standing, relaxed	1.2
Walking (on level surface)	
3.2 km/h	2.0
4.3 km/h	2.6
6.8 km/h	3.8
Office Activities	
Seated	1.0
Reading/writing/typing, seated	1.1
Filing, seated	1.2
Filing, standing	1.4
Walking about	1.7
Lifting/packing	2.1

References

[1] N.E. Klepeis, W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J.V. Behar, S.C. Hern, W.H. Engelmann, The national human activity pattern survey (nhaps): a resource for assessing exposure to environmental pollutants, J. Expo. Sci. Environ. Epidemiol. 11 (2001) 231–252.

[2] B.R. Kingma, A.J. Frijns, L. Schellen, W.D. van Marken Lichtenbelt, Beyond the classic thermoneutral zone, Temperature 1 (2014) 142–149, http://dx.doi.org/10.4161/temp.29702, pMID: 27583296.

[3] B. Kingma, M. Schweiker, A. Wagner, W.D. van Marken Lichtenbelt, Exploring internal body heat balance to understand thermal sensation, Build. Res. Inf. 45 (2017) 808–818, http://dx.doi.org/10.1080/09613218.2017.1299996.

[4] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (2008) 394–398, http://dx.doi.org/10.1016/j.enbuild.2007.03.007.

[5] P. Urban, W. Sven, Quantifying the heating and cooling demand in Europe, 2015.

[6] L. Pastore, M. Andersen, Building energy certification versus user satisfaction with the indoor environment: Findings from a multi-site post-occupancy evaluation (poe) in Switzerland, Build. Environ. 150 (2019) 60–74, http://dx.doi.org/10.1016/j.buildenv.2019.01.001.

[7] M. Frontczak, P. Wargocki, Literature survey on how different factors influence human comfort in indoor environments, Build. Environ. 46 (2011) 922–937, http://dx.doi.org/10.1016/j.buildenv.2010.10.021.

[8] M.A. Humphreys, J.F. Nicol, The validity of iso-pmv for predicting comfort votes in every-day thermal environments, Energy Build. 34 (2002) 667–684.

[9] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, L. Soibelman, One size does not fit all: Understanding user preferences for building automation systems, Energy Build. 145 (2017) http://dx.doi.org/10.1016/j.enbuild.2017.04.015.

[10] M. Schweiker, G.M. Huebner, B.R.M. Kingma, R. Kramer, H. Pallubinsky, Drivers of diversity in human thermal perception – a review for holistic comfort models, Temperature 5 (2018) 308–342, http://dx.doi.org/10.1080/23328940.2018.1534490, pMID: 30574525.

[11] Z. Wang, R. de Dear, M. Luo, B. Lin, Y. He, A. Ghahramani, Y. Zhu, Individual difference in thermal comfort: A literature review, Build. Environ. 138 (2018) 181–193, http://dx.doi.org/10.1016/j.buildenv.2018.04.040.

[12] D. Weinert, J. Waterhouse, The circadian rhythm of core temperature: effects of physical activity and aging, Physiol. Behav. 90 (2007) 246–256.

[13] A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo, Harrison's Principles of Internal Medicine, twentieth ed., McGraw-Hill Education, 2018.

[14] H. Zhang, E. Arens, C. Huizenga, T. Han, Thermal sensation and comfort models for non-uniform and transient environments, part ii: Local comfort of individual body parts, Build. Environ. 45 (2010a) 389–398.

[15] H. Zhang, E. Arens, C. Huizenga, T. Han, Thermal sensation and comfort models for non-uniform and transient environments: Part i: Local sensation of individual body parts, Build. Environ. 45 (2010b) 380–388.

[16] J.-H. Choi, V. Loftness, Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations, Build. Environ. 58 (2012) 258–269.

[17] C.F. Bulcao, S.M. Frank, S.N. Raja, K.M. Tran, D.S. Goldstein, Relative contribution of core and skin temperatures to thermal comfort in humans, J. Therm. Biol. 25 (2000) 147–150.

[18] A.A. Romanovsky, Skin temperature: its role in thermoregulation, Acta Physiol. 210 (2014) 498–507.

[19] G. Havenith, Interaction of clothing and thermoregulation, Exog. Dermatol. 1 (2002) 221–230.

[20] R. Rawal, M. Schweiker, O.B. Kazanci, V. Vardhan, Q. Jin, L. Duanmu, Personal comfort systems: A review on comfort, energy, and economics, Energy Build. 214 (2020) 109858.

[21] H. Zhang, E. Arens, Y. Zhai, A review of the corrective power of personal comfort systems in non-neutral ambient environments, Build. Environ. 91 (2015) 15–41.

[22] K. Chen, Q. Xu, B. Leow, A. Ghahramani, Personal thermal comfort models based on physiological measurements–a design of experiments based review, Build. Environ. (2022) 109919.

[23] Y. Takahashi, A. Nomoto, S. Yoda, R. Hisayama, M. Ogata, Y. Ozeki, S. i. Tanabe, Thermoregulation model jos-3 with new open source code, Energy Build. 231 (2021) 110575.

[24] J.A. Stolwijk, A mathematical model of physiological temperature regulation in man, 1971.

[25] M. Fu, W. Weng, W. Chen, N. Luo, Review on modeling heat transfer and thermoregulatory responses in human body, J. Therm. Biol. 62 (2016) 189–200.

[26] D. Fiala, K.J. Lomas, M. Stohrer, A computer model of human thermoregulation for a wide range of environmental conditions: the passive system, J. Appl. Physiol. 87 (1999) 1957–1972.

[27] C. Huizenga, Z. Hui, E. Arens, A model of human physiology and comfort for assessing complex thermal environments, Build. Environ. 36 (2001) 691–699.

[28] W. Karaki, N. Ghaddar, K. Ghali, K. Kuklane, I. Holmér, L. Vanggaard, Human thermal response with improved ava modeling of the digits, Int. J. Therm. Sci. 67 (2013) 41–52.

[29] L. Schellen, M. Loomans, B. Kingma, M. De Wit, A. Frijns, W. van Marken Lichtenbelt, The use of a thermophysiological model in the built environment to predict thermal sensation: coupling with the indoor environment and thermal sensation, Build. Environ. 59 (2013) 10–22.

[30] B. Kingma, Human Thermoregulation: A Synergy Between Physiology and Mathematical Modelling (Ph.D. dissertation), Maastricht University, 2012.

[31] M. Rida, A. Frijns, D. Khovalyg, Modeling local thermal responses of individuals: Validation of advanced human thermo-physiology models, Build. Environ. 243 (2023) 110667, http://dx.doi.org/10.1016/j.buildenv.2023.110667.

[32] D. Khovalyg, Y. Ravussin, Interindividual variability of human thermoregulation: Toward personalized ergonomics of the indoor thermal environment, Obesity 30 (2022) 1345–1350, http://dx.doi.org/10.1002/oby.23454.

[33] E. Arens, H. Zhang, C. Huizenga, Partial-and whole-body thermal sensation and comfort—part i: Uniform environmental conditions, J. Therm. Biol. 31 (2006) 53–59.

[34] J. Kim, S. Schiavon, G. Brager, Personal comfort models–a new paradigm in thermal comfort for occupant-centric environmental control, Build. Environ. 132 (2018) 114–124.

[35] S. Liu, S. Schiavon, H.P. Das, M. Jin, C.J. Spanos, Personal thermal comfort models with wearable sensors, Build. Environ. 162 (2019) 106281.

[36] Z. Wang, R. Matsuhashi, H. Onodera, Towards wearable thermal comfort assessment framework by analysis of heart rate variability, Build. Environ. 223 (2022) 109504.

[37] T. Chaudhuri, Y.C. Soh, H. Li, L. Xie, Machine learning based prediction of thermal comfort in buildings of equatorial singapore, in: 2017 IEEE International Conference on Smart Grid and Smart Cities, ICSGSC, IEEE, 2017, pp. 72–77.

[38] B. Yang, X. Li, Y. Hou, A. Meier, X. Cheng, J.-H. Choi, F. Wang, H. Wang, A. Wagner, D. Yan, et al., Non-invasive (non-contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-a review, Energy Build. (2020) 110261.

[39] C. Yu, B. Li, Y. Wu, B. Chen, R. Kosonen, S. Kilpelainen, H. Liu, Performances of machine learning algorithms for individual thermal comfort prediction based on data from professional and practical settings, J. Build. Eng. 61 (2022) 105278.

[40] J. Lyu, H. Du, Z. Zhao, Y. Shi, B. Wang, Z. Lian, Where should the thermal image sensor of a smart a/c look?—occupant thermal sensation model based on thermal imaging data, Build. Environ. (2023) 110405.

[41] W. Li, J. Zhang, T. Zhao, R. Liang, Experimental research of online monitoring and evaluation method of human thermal sensation in different active states based on wristband device, Energy Build. 173 (2018) 613–622.

[42] A. Ghahramani, G. Castro, S.A. Karvigh, B. Becerik-Gerber, Towards unsupervised learning of thermal comfort using infrared thermography, Appl. Energy 211 (2018) 41–49.

[43] F. Jazizadeh, W. Jung, Personalized thermal comfort inference using rgb video images for distributed hvac control, Appl. Energy 220 (2018) 829–841.

[44] D. Li, C.C. Menassa, V.R. Kamat, Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras, Appl. Energy 251 (2019) 113336.

[45] A. Aryal, B. Becerik-Gerber, Thermal comfort modeling when personalized comfort systems are in use: Comparison of sensing and learning methods, Build. Environ. 185 (2020) 107316.

[46] A. Aryal, B. Becerik-Gerber, A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor, Build. Environ. 160 (2019) 106223.

[47] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2d pose estimation using part affinity fields, in: Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 7291–7299.
[48] G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, K. Murphy, Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model, in: European Conference on Computer Vision, ECCV, 2018, pp. 269–286.
[49] S. Kreiss, L. Bertoni, A. Alahi, Pifpaf: Composite fields for human pose estimation, in: Conference on Computer Vision and Pattern Recognition, CVPR, 2019.
[50] Y. Xu, J. Zhang, Q. Zhang, D. Tao, Vitpose: Simple vision transformer baselines for human pose estimation, 2022, arXiv preprint arXiv:2204.12484.
[51] T. Kim, Generalizing mlps with dropouts, batch normalization, and skip connections, 2021, arXiv preprint arXiv:2108.08186.
[52] E. Eidinger, R. Enbar, T. Hassner, Age and gender estimation of unfiltered faces, IEEE Trans. Inf. Forensics Secur. 9 (2014) 2170–2179.
[53] R. Rothe, R. Timofte, L. Van Gool, Dex: Deep expectation of apparent age from a single image, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 10–15.
[54] H. Xiao, K. Rasul, R. Vollgraf, Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017, arXiv preprint arXiv:1708.07747.
[55] Z. Liu, P. Luo, S. Qiu, X. Wang, X. Tang, Deepfashion: Powering robust clothes recognition and retrieval with rich annotations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1096–1104.
[56] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022, arXiv preprint arXiv:2207.02696.
[57] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, et al., Internimage: Exploring large-scale vision foundation models with deformable convolutions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 14408–14419.
[58] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September (2014) 6-12, Proceedings, Part V 13, Springer, 2014, pp. 740–755.
[59] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, J. Sun, Crowdhuman: A benchmark for detecting human in a crowd, 2018, arXiv preprint arXiv:1805.00123.
[60] D. Hoiem, S.K. Divvala, J.H. Hays, Pascal voc 2008 challenge, World Lit. Today 24 (2009).
[61] R. Wang, D. Chen, Z. Wu, Y. Chen, X. Dai, M. Liu, L. Yuan, Y.-G. Jiang, Masked video distillation: Rethinking masked feature modeling for self-supervised video representation learning, 2022, arXiv preprint arXiv:2212.04500.
[62] L. Wang, B. Huang, Z. Zhao, Z. Tong, Y. He, Y. Wang, Y. Wang, Y. Qiao, Videomae v2: Scaling video masked autoencoders with dual masking, 2023, arXiv preprint arXiv:2303.16727.
[63] C. Feichtenhofer, H. Fan, J. Malik, K. He, Slowfast networks for video recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6202–6211.
[64] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, et al., The kinetics human action video dataset, 2017, arXiv preprint arXiv:1705.06950.
[65] J. Deng, J. Guo, E. Ververas, I. Kotsia, S. Zafeiriou, Retinaface: single-shot multi-level face localisation in the wild, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5203–5212.
[66] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4690–4699.
[67] Y. Ge, R. Zhang, X. Wang, X. Tang, P. Luo, Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5337–5345.
[68] I. O. f. S. ISO7730, Ergonomics of the Thermal Environment: Analytical Determination and Interpretation of Thermal Comfort using Calculation of the PMV and PPD Indices and Local Thermal Comfort Criteria, International Organization for Standardization, 2005.
[69] S. Kreiss, L. Bertoni, A. Alahi, Openpifpaf: Composite fields for semantic keypoint detection and spatio-temporal association, IEEE Trans. Intell. Transp. Syst. 23 (2021) 13498–13511.
[70] A. Sellers, D. Khovalyg, W. van Marken Lichtenbelt, Thermoregulation of tuvan pastoralists and western europeans during cold exposure, Am. J. Hum. Biol. (2023) e23933, http://dx.doi.org/10.1002/ajhb.23933.
[71] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[72] G. Havenith, R. Heus, W. Lotens, Resultant clothing insulation: a function of body movement, posture, wind, clothing fit and ensemble thickness, Ergonomics 33 (1990) 67–84.
[73] G. Havenith, K. Kuklane, J. Fan, S. Hodder, Y. Ouzzahra, K. Lundgren, Y. Au, D. Loveday, A database of static clothing thermal insulation and vapor permeability values of non-western ensembles for use in ashrae standard 55, iso 7730, and iso 9920, Ashrae Trans. 121 (2015) 197–215.
[74] C.E. Smith, A. transient, Three-Dimensional Model of the Human Thermal System, Kansas State University, 1991.
[75] N. Pavlichenko, D. Ustalov, Imdb-wiki-sbs: An evaluation dataset for crowdsourced pairwise comparisons, 2021, arXiv preprint arXiv:2110.14990.
[76] G. Levi, T. Hassner, Age and gender classification using convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 34–42.
[77] J. Carreira, E. Noland, C. Hillier, A. Zisserman, A short note on the kinetics-700 human action dataset, 2019, arXiv preprint arXiv:1907.06987.
[78] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.
[79] D. Zauss, S. Kreiss, A. Alahi, Keypoint communities, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 11057–11066.
[80] ASHRAE, Ashrae 55-2010: Thermal environmental conditions for human occupancy, 2010.