



## OPEN Deep learning-based infrared thermography reveals reproducible uniform and individual thermoregulatory responses during running

Vincent Weber<sup>1</sup>, Daniel Andrés López<sup>2</sup>, David Tobias Ochmann<sup>1</sup>, Severin Zentgraf<sup>1</sup>, Markus Nägele<sup>3</sup>, Elmo W. I. Neuberger<sup>1</sup>, Elmar Schömer<sup>2</sup>, Perikles Simon<sup>1</sup> & Barlo Hillen<sup>1,4</sup>✉

Infrared thermography (IRT) has recently gained attention in the field of exercise physiology, due to its ability to monitor thermoregulatory and cardiopulmonary responses non-invasively and in real time during physical exercise. However, the reproducibility of intra-individual measurement and standardization of region-of-interest selection in relation to the acute exercise response remain inconclusive. This study aimed to examine the reproducibility and physiological relevance of specific skin temperature ( $T_{SK}$ ) metrics processed automatically using deep learning-assisted IRT during running, and to synchronize these metrics with cardiopulmonary and thermoregulatory parameters. Eleven endurance-trained individuals performed three 46-min running sessions over 2 days, with the same average external load but different intensity distributions. Individual anaerobic threshold velocity (vIAT), previously determined by cardiopulmonary exercise testing, was used to prescribe running intensity. During exercise, oxygen consumption ( $VO_2$ ), core temperature ( $T_{CORE}$ ), heart rate (HR) and different  $T_{SK}$  metrics, including non-vessel ( $T_{NV}$ ), cutaneous arterial perforator ( $T_p$ ), and superficial vein patterns, were continuously measured. All  $T_{SK}$  metrics displayed consistent temporal dynamics aligned with external load, but their absolute temperature levels differed systematically. During intermittent running and recovery,  $T_p$  exhibited robust correlations with HR and  $VO_2$  ( $r = -0.63$  to  $-0.9$ ,  $p < 0.001$ ), and  $T_p$  entropy showed consistent associations with  $T_{CORE}$  during the warm-up ( $r = 0.59-0.83$ ,  $p < 0.001$ ). This indicates uniform response patterns across the cohort. In contrast,  $T_{NV}$  demonstrated heterogeneous correlations with  $T_{CORE}$  depending on individual exercise capacity. A strong inverse correlation was identified between  $\Delta T_{NV}$  and vIAT ( $r = -0.74$  to  $-0.88$ ,  $p \leq 0.009$ ) and individuals with higher vIAT demonstrated greater  $T_{CORE}-T_{NV}$  gradients during running. Measurements of  $\Delta T_{NV}$  demonstrated high reproducibility, with intra-individual ICC(3,1) values of 0.89 for recovery and 0.76 for warm-up, and no statistically significant differences between the three sessions. Deep learning-assisted IRT provides reproducible, physiologically consistent metrics across repeated exercise sessions, regardless of the day or prior load. Distinct  $T_{SK}$  metrics capture both uniform and individual-specific thermoregulatory responses. Variability in peripheral temperature regulation is more strongly associated with running velocity at the individual anaerobic threshold than with maximal cardiorespiratory fitness.

**Keywords** Infrared thermal imaging, Thermography, Core body temperature, Reproducibility, Exercise physiology, Deep neural network, Artificial intelligence

<sup>1</sup>Department of Sports Medicine, Disease Prevention and Rehabilitation, Institute of Sports Science, Faculty of Social Science, Media and Sports, Johannes Gutenberg University, Mainz, Germany. <sup>2</sup>Institute of Computer Science, Research Group Computational Geometry, Johannes Gutenberg University, Mainz, Germany. <sup>3</sup>Optoprecision GmbH, Bremen, Germany. <sup>4</sup>Institute of Occupational, Social, and Environmental Medicine, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany. ✉email: b.hillen@uni-mainz.de

The application of infrared thermography (IRT) during physical exercise is promising because it enables non-contact detection of acute cardiovascular and thermoregulatory responses<sup>1,2</sup> based on exercise-induced variations in skin temperature ( $T_{SK}$ )<sup>3</sup>. Furthermore, given the interplay between core temperature ( $T_{CORE}$ ) and  $T_{SK}$ , for example, through vascular reactivity,  $T_{SK}$  measurements can support diagnostics in cardiovascular health<sup>4</sup>, and potentially serve as a reliable proxy for thermoregulation, helping to assess the risk of exertional heat illness<sup>2</sup>. However, the literature on influential factors and intra-individual reproducibility on IRT measurements during exercise remains scarce.

Applying IRT in the context of endurance exercise may provide real-time insights into athletes' internal load<sup>2,5-9</sup>. Notably, the thermal pattern of perforator vessels appears across the body during endurance exercises and varies locally. Recently, this pattern has been proposed to be of particular interest in field test scenarios<sup>10</sup>. In addition, thermal pattern distribution and entropy analysis in the chest region have been demonstrated to be associated with cardiorespiratory fitness<sup>11</sup>. Interestingly, this pattern was already analysed in 1992 but was not further investigated in subsequent studies<sup>12</sup>.

In running, Belinchón-deMiguel et al.<sup>13</sup> reported that  $T_{SK}$  variations detectable by IRT are related to physiological and biomechanical load during an ultra-endurance race. Similarly, another study found that IRT captures acute metabolic stress responses during racewalking and provides insights into individual physiological adaptation<sup>7</sup>. In rowing, Silva et al.<sup>14</sup> demonstrated that high-intensity rowing induces distinct thermal responses in the upper and lower body. Increased temperature in active muscle groups reflects elevated metabolic activity and blood flow, whereas cooler skin regions indicate reduced perfusion and muscle recruitment. Overall, thermal pattern distributions are influenced by thermoregulatory mechanisms such as sweating and blood redistribution. In cycling, Kapoor et al.<sup>15</sup> and Jastrzębska et al.<sup>16</sup> reported significant correlations between aerobic fitness and  $T_{SK}$  variations. Further studies have demonstrated that thermal pattern distribution and entropy analysis of thermal images can be used to assess physiological stress responses during cycling<sup>11,17</sup>. Novotny et al.<sup>18</sup> showed that specific muscle activity patterns in breaststroke swimming can be visualised using IRT, indicating that IRT can be utilized for biomechanical analysis in swimmers. In addition, IRT was able to detect differences in thermal responses to endurance exercise in cross-country skiers and swimmers, which were associated with sport-specific physiological demands<sup>6</sup>.

The current scientific debate on applying IRT in endurance performance diagnostics focuses on whether it can detect acute physiological responses and on the influence of factors, such as sweat rate<sup>19</sup>,  $T_{CORE}$ <sup>20</sup>, ambient temperature<sup>21</sup>, region of interest (ROI)<sup>1,22,23</sup>, exercise intensity<sup>3</sup>, circadian rhythm<sup>24</sup>, and subcutaneous fat<sup>25,26</sup>.  $T_{CORE}$  in particular is contested, both with respect to the use of IRT as a non-invasive indicator of  $T_{CORE}$  and regarding the influence of  $T_{CORE}$  on  $T_{SK}$  variations during exercise<sup>1,20,27,28</sup>. It has been proposed that specific ROIs are not suitable for indicating changes in  $T_{CORE}$ <sup>1,27</sup>. Furthermore, although no significant correlations between  $T_{CORE}$  and  $T_{SK}$  variations have been observed<sup>28</sup>, the core-to-skin temperature gradient is a physiologically relevant marker of thermoregulatory status that is affected by aerobic capacity<sup>29</sup>. This gradient is usually assessed using skin-mounted temperature sensors<sup>29-31</sup>, and IRT-based evaluation has so far been limited to the clinical setting of septic shock<sup>32</sup>. However, measurements of  $T_{CORE}$  variations under thermoneutral conditions is frequently not included in IRT experiments, highlighting the need for further research in this area.

In summary, the available findings are promising, but the research landscape in the field is characterized by heterogeneous study designs, which reduces the comparability of results and hinders standardization. In addition, investigations into the use of IRT across different endurance exercise protocols and in retest scenarios are limited, thereby impeding the validation of associations between  $T_{SK}$  variations and physiological traits, and intra-individual reproducibility<sup>23</sup>.

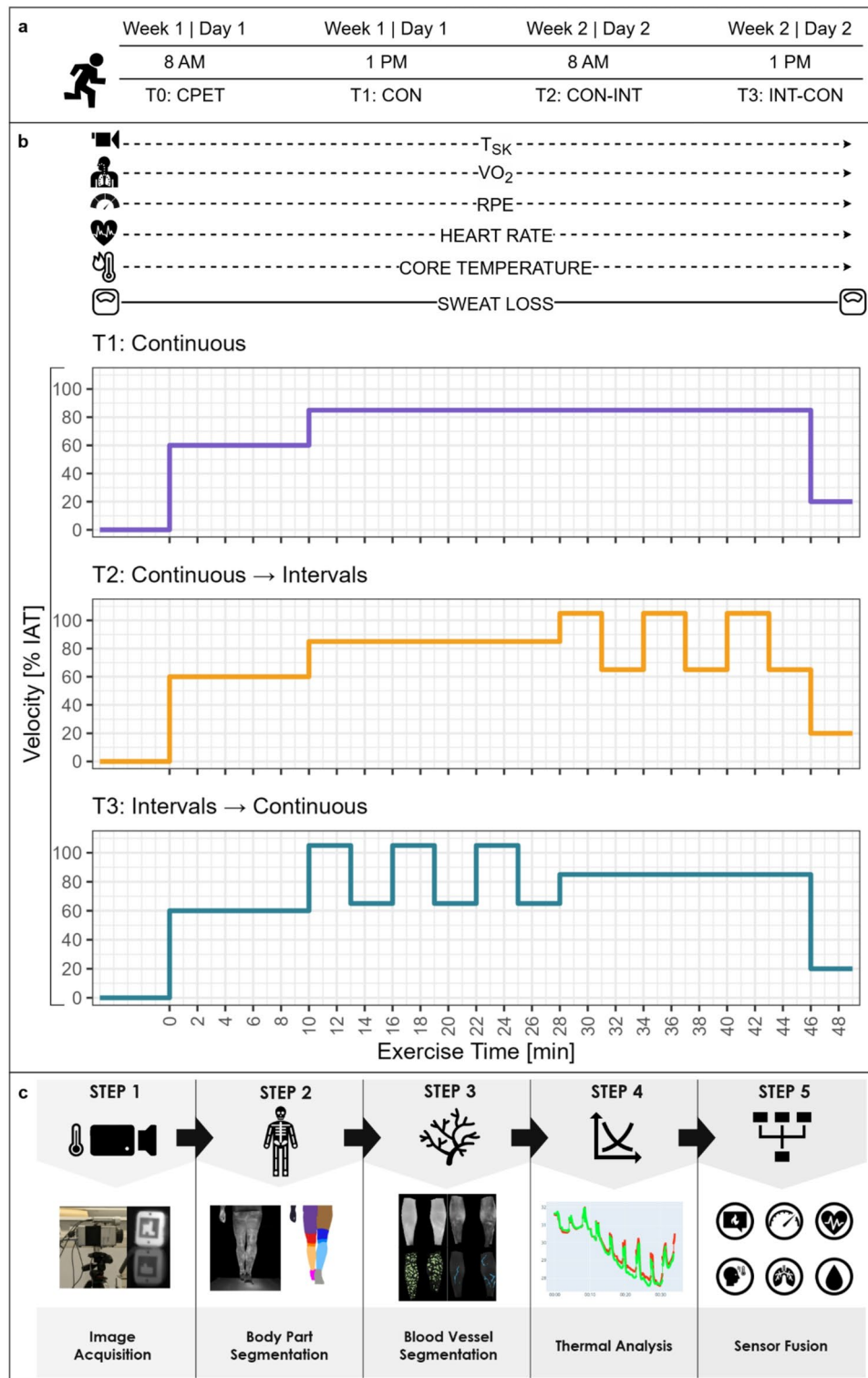
Recently, advanced deep learning-based automatic analysis of thermograms of the posterior legs captured during running has been demonstrated, enabling objective and reproducible time series analysis of several statistical variables in specific ROIs to comprehensively explore synchronized cardiopulmonary and thermographic data<sup>33-35</sup>. Therefore, IRT coupled with artificial intelligence-driven automatic analysis is required to provide further evidence for preliminary assumptions and to enable objective thermogram analysis. Additionally, the calves appear to be a preferable, practical, and promising ROI in running and cycling IRT experiments<sup>1,36</sup>.

In the context of the current state of research, three exploratory research objectives were defined for this study. The aim was to explore: (i) the association of distinct  $T_{SK}$  metric variations with cardiopulmonary responses during different running protocols with a similar average external load, (ii) the potential physiological and environmental factors that influence  $T_{SK}$  variations and (iii) the reproducibility of inter- and intra-individual  $T_{SK}$  variations under thermoneutral conditions.

## Methods

### Study design and participants

Participant recruitment and examinations were conducted between March and June 2022. The participants performed three running sessions and one incremental cardiopulmonary exercise test (CPET) on a treadmill in the laboratory of the Department of Sports Medicine, Prevention and Rehabilitation at Johannes Gutenberg-University in Mainz, Germany. In the first week, they performed the initial CPET (T0) to exclude exercise-induced cardiovascular risks and to determine the external load for the subsequent three submaximal running sessions. In the afternoon of the same day, they performed a continuous running session (T1). In the following week, participants performed their third (morning, T2) and fourth (afternoon, T3) running sessions, consisting of continuous and intermittent load within a single day. During all experiments, cardiopulmonary, metabolic, and thermo-physiological parameters were monitored synchronously and continuously (Fig. 1). Healthy individuals aged 18–30 years with a physical activity rating of at least 6 points were eligible for inclusion. Individuals with any of the following characteristics were excluded: inability to perform exercise, acute injury, extensive scars



**Fig. 1.** Overview of the study design and data processing pipeline. **(a)** Timeline of the baseline cardiopulmonary exercise test (CPET) and the three exercise sessions (T1, T2 and T3) with continuous (CON) and intermittent (INT) external load. **(b)** Overview of the three running protocols and the measured parameters during the exercise sessions. The different skin temperature ( $T_{SK}$ ) metrics, oxygen consumption ( $VO_2$ ), rate of perceived exertion (RPE), heart rate and core temperature were continuously measured throughout the exercise (dashed arrows) and sweat loss was measured before and after the exercise (solid line). **(c)** IRT data processing pipeline from Andrés López et al.<sup>35</sup>.

or tattoos on relevant ROIs, body mass < 40 kg, intestinal disease, increased risk of intestinal disease, motility disorders of the gastrointestinal tract, surgical interventions in the gastrointestinal tract, swallowing disorders (e.g., gag reflex), chronic disease, use of acute or chronic medication, pacemaker or electromedical implants, or pregnancy. By informing the participants in advance and using an additional medical history form prior to the examination, the participation of individuals who did not meet the criteria was prevented.

Sixteen endurance-trained individuals who self-reported being healthy at the time of recruitment were enrolled in the experiment. Twelve of the sixteen completed all running sessions. Four individuals were excluded after CPET (T0) due to medical issues (e.g., suspected post covid-19 condition), and one individual was removed prior to data analysis due to irregular non-uniformity temperature drift correction. Consequently, the final analytical sample comprised 11 healthy participants without any diagnosed or identified medical conditions. All procedures performed were reviewed by the Ethics Committee of the State Medical Association of Rhineland-Palatinate and complied with the Declaration of Helsinki (ethical approval identifier: 2021–15713). Additionally, this exploratory single-arm study was registered on 23 June 2022 in the German Clinical Trial Register (clinical trial number: DRKS00029114; URL: <https://drks.de/search/en/trial/DRKS00029114>). All participants were informed about the study procedures and provided written informed consent.

### Baseline assessment and cardiopulmonary exercise testing

At baseline, participants rated their habitual physical activity (PA-R)<sup>37</sup>, underwent a resting pulmonary function test (Bodybox 5500, MEDISOFT GmbH, Hamburg, Germany), and a bioimpedance analysis (InBody 3.0, Biospace, Seoul, Korea). Afterwards, they performed a stepwise incremental CPET on a treadmill (Saturn, HP cosmos, Nussdorf-Traunstein, Germany). The CPET protocol started after a 10-min acclimatisation period at 6 km/h with an incline of 1.5%. After every 3-min stage, separated by a short resting break (45 s), the velocity was increased by 2 km/h until maximum voluntary exhaustion. During the CPET, respiration was measured by breath-by-breath analysis (Ergostik, Blue Cherry, Geratherm Respiratory GmbH, Bad Kissingen, Germany), heart rate (HR) was recorded with a chest-strap monitor (Polar H10, Polar Electro, Kempele, Finland), and rate of perceived exertion (RPE) was assessed using the BORG scale [6–20]<sup>38</sup>. Capillary blood samples were obtained from the earlobe at rest, after the end of each intensity stage, and at volitional exhaustion. Blood lactate concentration was analysed immediately on site (EKF-diagnostic GmbH, Magdeburg, Germany), and the individual anaerobic threshold (IAT) was determined according to Dickhuth's model<sup>39</sup>. Adverse events occurring during the examination procedures were monitored and recorded for all participants.

### Three different running protocols with the same average external load

Each running protocol began with a 10-min acclimatisation period in a standing resting position on the treadmill, followed by a 10-min warm-up period with an external load equivalent to 60% of the running velocity at the IAT (vIAT). The incline of the treadmill was set to 1.5% for each running session. After the warm-up, the first running protocol (T1) continued for an additional 36 min at a constant external load (CON) corresponding to 85% of the vIAT and was performed on the first experimental day between 1:00 and 2:30 PM. The second running protocol (T2) consisted of an continuous external load from min 10 to min 28, followed by an intermittent load (INT) from min 28 to min 46. The intermittent part comprised three sequences of 3 min at 105% of vIAT alternating with 3 min at 65% of vIAT. Participants performed this protocol on the second experimental day between 8:00 and 9:30 AM. The third running session (T3) was also conducted on the second experimental day, between 1:00 and 2:30 PM. This protocol included the same 18-min intermittent and continuous running periods, but in reverse order. Each running protocol concluded with a 3-min recovery period of walking at a velocity of 4 km/h. The total duration of each running session was 49 min. In addition to cardiopulmonary parameters,  $T_{\text{CORE}}$  was measured using ingestible telemetric pills (Telemetric System for Continuous Gastrointestinal Temperature Monitoring, Herouville ST Claire France) on the second experimental day, and non-invasive  $T_{\text{CORE}}$  ANT+ sensors (CORE 1 Sensor, greenteg AG, Zürich, Switzerland) on both experimental days. The telemetric pills were ingested 1 h prior to the first exercise session, as Notley et al.<sup>40</sup> demonstrated no significant differences in core temperature recordings between pills ingested 12, 6, 3–1 h before data collection across rest, exercise and recovery conditions. To avoid confounding temperature measurements, participants were not permitted to consume any food or fluids between the ingestion of the telemetric capsule and the termination of the first exercise session. The individual metabolic heat production rate (MHP [ $\text{W}/\text{m}^2$ ]) was calculated using the metabolic rate obtained through indirect calorimetry, body surface area (BSA [ $\text{m}^2$ ]), and the external work rate [ $\text{W}$ ]<sup>41</sup>. BSA was estimated using Mosteller's formula<sup>42</sup> from the individual body mass and body height. The quantity of sweat lost was calculated as the difference in body mass before and after exercise. Participants were instructed to void their bladders prior to being weighed, in minimal clothing, using calibrated digital scales accurate to 0.1 kg, both immediately before and after the exercise session. No fluids were consumed, and no urine was voided during the exercise session to ensure that the change in body mass accurately reflected sweat loss. An overview of the running protocols, including the measured parameters, is provided in Fig. 1.

### Infrared thermography protocol

The thermal camera specification and calibration routine are described in Andrés López et al.<sup>35</sup> (VarioCam HD head 800, JENOPTIK AG, Jena, Germany; detector type: uncooled microbolometer with  $1024 \times 768$  IR pixels, long-wave infrared 7.5–14  $\mu\text{m}$ , 30 frames per second (fps) with rolling shutter; thermal resolution 0.02 K, measurement accuracy  $\pm 1.0$  K). The camera was placed 235 cm behind the ROI, and the optical axis of the camera is perpendicular to the ROI surface. The temperature range was set to 25–35 °C and the emissivity to 0.98 by the external two-point calibration device, allowing each thermogram to be consistently calibrated with temperature drift correction. Sources of external heat radiation were minimized by keeping windows and doors closed, blinds lowered, and the number of people in the room was kept constant to avoid influencing  $T_{\text{SK}}$  by drafts

and to maintain stable laboratory conditions. Room temperature [°C] and humidity [%] were measured using a hygrometer (Klimalogg Pro Thermo-Hygro-Station, TFA-Germany, Wertheim-Reicholzheim, accuracy  $\pm 1$  °C and  $\pm 3\%$  relative humidity). Before starting the measurements, the laboratory was cooled down to temperatures ranging between  $19.7 \pm 0.97$  °C and  $20.7 \pm 0.89$  °C, with humidity levels between  $42 \pm 5.1\%$  and  $45.3 \pm 6.3\%$ . This cooling process was undertaken to ensure thermoneutral conditions, with the objective of focusing on exercise-induced thermoregulatory responses while minimizing the impact of confounding environmental thermal stress. This approach is consistent with the majority of studies examining exercise-induced skin temperature changes<sup>2</sup>. Prior to the measurement days, participants followed a strict preparation protocol to standardise external and internal influential factors that could affect the procedures. Strenuous physical activity and the use of skin care and shaving products were prohibited in the 24 h before the experiments. Further instructions included not showering on the morning of testing; abstaining from food and drink for 2 h before the experiment; limiting fluid intake to  $\leq 2$  L in the preceding 10 h; avoiding caffeinated beverages and alcohol; and obtaining at least 8 h of sleep the night before. Travel to the laboratory should minimize physical exertion (car or public transport). During testing, participants wore standardized clothing: running shoes, ankle socks, knee-length (or shorter) running shorts and, where applicable, a sports bra; long hair was to be secured with a hair tie.

### Deep learning-assisted automatic thermogram analysis

The processing pipeline by Andrés López et al.<sup>43</sup>, from acquisition to deep neural network (DNN) analysis, executed ROI extraction of the whole calf on both legs followed by the identification of blood vessel-related patterns of superficial veins and cutaneous arterial perforator vessels. As a further development of previous work, the semantic segmentation models were optimized using 670 training images and were evaluated using 200 test images for both model types—body parts and blood vessels. We applied a DeepLabv3 +<sup>44</sup> model for the body part network, with the following hyperparameters: regional mutual information loss<sup>45</sup>, a learning rate of 0.000686, and a batch size of 8 optimized using the AdamW algorithm<sup>46</sup>. The vessel network was implemented using an Attention-U-Net<sup>47</sup> architecture, a Tanimoto loss function<sup>48</sup>, a learning rate of 0.000052 and an AdaBelief<sup>49</sup> optimization algorithm. Finally, four different  $T_{SK}$  metrics were calculated: mean surface radiation temperature ( $T_{MEAN}$ ), non-vessel surface radiation temperature ( $T_{NV}$ ), perforator surface radiation temperature ( $T_p$ ) and vein surface radiation temperature ( $T_v$ ), and their associated Shannon entropy of pixel intensities<sup>50</sup>.

### Integration of IRT and sensors

For comparative analysis, data from various measurement systems, including IRT, HR monitors, ingestible  $T_{CORE}$  pills, ANT+  $T_{CORE}$  sensors, metabolic analyzers, and environmental data, were combined into a single dataset. Although the IRT data provides 30 fps, not all thermograms offer the same level of insight. The analysis focused on the stance phase of a single step, when the leg is in extension. The swing (flight) phase was excluded due to substantial motion blur and frequent occlusions of the ROI. Valid images were selected individually by identifying the largest segmented calf in a series of images and treating these images separately. The low-frequency sensor data were then resampled to align with the timing of the high-frequency IRT data. If necessary, a Savitzky–Golay filter with a window length of 151 was applied to smooth the IRT and sensory data and reduce high-frequency noise. To provide an overview of a single experiment, each stage was represented by a single data row corresponding to the mean value over the final five seconds of the stage. The acceleration and deceleration phases of the treadmill were excluded. Data integration was performed using python (v3.8.13), pandas (v2.0.3), OpenCV (v4.8.1), and SciPy (v1.10.1). DNN training was carried out with torch (v2.4.1) and lightning (v2.3.3) on a CUDA-enabled device.

### Statistical analysis

All statistical analyses and visualizations were conducted using R (v4.4.3) and the integrated development environment RStudio (v2025.05.1 + 513 for Windows). Numeric variables are generally presented as mean  $\pm$  standard deviation (SD), and categorical variables are expressed as n (%). All visualizations were generated using the *ggplot2* package (v3.5.1), except for the heatmaps which were created with the *ComplexHeatmap* package (v2.22.0). Data exclusions were applied due to technical issues or missing data points. The following combinations of participant, running session and phases were excluded from the heart rate specific analyses: P9/T1 (all phases), P1/T2 (second half), P6/T2 (second half), P1/T3 (second half) and P4/T3 (second half). For  $VO_2$ , the first half, second half, and recovery phases of P10/T2 were excluded. Core temperature data from the ANT+ sensor were excluded for P9/T1 (all phases), and ingested pill sensor data were excluded for P4/T2 (all phases). The statistical significance level was set at  $p < 0.05$ .

To visualize the mean  $T_{SK}$  metrics for all participants a 3rd-order digital low-pass Butterworth filter was applied using the *signal* package (v1.8.1) to suppress high-frequency noise, with a cutoff frequency of 0.04 Hz normalized by the Nyquist frequency. First, the relevant variables were aggregated by taking the median across participants at each timepoint. Subsequently, a rolling median filter with a window width of 350 samples was applied to further reduce noise. For a specific variable ( $T_v$ ), missing values were first linearly interpolated using the *zoo* package (v1.8.12) prior to filtering. To investigate the associations between temperature and physiological variables across multiple stages, we performed a systematic correlation analysis using parametric and non-parametric approaches, depending on the data distribution. Normality was assessed using the Shapiro–Wilk test with the *stats* package (v4.4.3). Pearson or Spearman rank correlation were then applied using the *stats* package as appropriate. When multiple data points from the same measurement were included in the analysis, the *rmcorr* package (v0.7.0) was used to account for variability in the intercepts across individuals. We report the repeated-measures correlation coefficient ( $r_{rm}$ ), the corresponding 95% confidence intervals (95% CI, via bootstrapping), and the significance levels. Correlation strength was interpreted according to Cohen's conventions, with  $r < 0.1$  trivial, 0.1–0.29 small, 0.3–0.49 moderate and  $\geq 0.5$  large<sup>51</sup>.

We fitted linear mixed-effects models using the *lmerTest* package (v3.1.3) to examine the effects of experimental conditions on  $T_{SK}$  metrics,  $T_{CORE}$  and RPE and to identify differences in  $\Delta T_{NV}$  across the different running sessions. Participant was included as random intercept to account for repeated observations within individuals. Fixed-effects structures followed the a priori design, including the relevant within- and between-subject factors and their necessary interactions. When factors had more than two levels, pairwise comparisons were obtained with the *emmeans* package (v1.11.0) using Tukey's honest significance difference adjustment for multiple testing. Model assumptions were evaluated using the *performance* (v0.15.2) to generate residual-versus-fitted, Q-Q plots, and related diagnostic plots, and were complemented with simulation-based residuals checks from *DHARMa* (v0.4.7), including formal tests for outliers and dispersion. If diagnostics indicated violations such as non-normal residuals, heteroscedasticity, or influential outliers, we refitted the model with *robustlmm* (v3.3.1) to obtain robust estimates and confirmed that the overall pattern of fixed-effect results was stable across the two approaches.

We calculated the intra-class correlation (ICC) using the *irr* package (v0.84.1). Specifically, we used a two-way mixed-effects model (model = "twoway") to estimate ICC based on the chosen ICC type (e.g., "consistency" or "agreement") and calculated the reliability of single measurements (unit = "single"), which is suitable when individual observations rather than average scores are of interest. ICCs were calculated for each individual during six different exercise sections [T1-T2-T3 (warm-up), T1-T2 (CON), T1-T3 (CON), T2-T3 (CON), T2-T3 (INT), T1-T2-T3 (recovery)]. The interpretation of ICC values followed conventional thresholds: values below 0.5 indicated poor reliability, values between 0.5 and 0.75 indicated moderate reliability, values between 0.75 and 0.9 indicated good reliability, and values above 0.9 indicated excellent reliability<sup>52</sup>.

## Results

All individuals included were of normal weight, with a BMI ranging from 20.5 to 24.3 kg/m<sup>2</sup>, and a body fat percentage ranging from 10.2 to 18.1% for male participants, and from 13.6 to 25.3% for female participants.

The participants were well-trained, characterized by  $VO_{2max}$  between 41 and 65.9 mL/min/kg and vIAT between 9.9 and 16.5 km/h, and were engaged in intense exercise on a regular basis, as indicated by their PA-R scores ranging from 6 to 10 points (Table 1). No adverse events were observed during the study procedure.

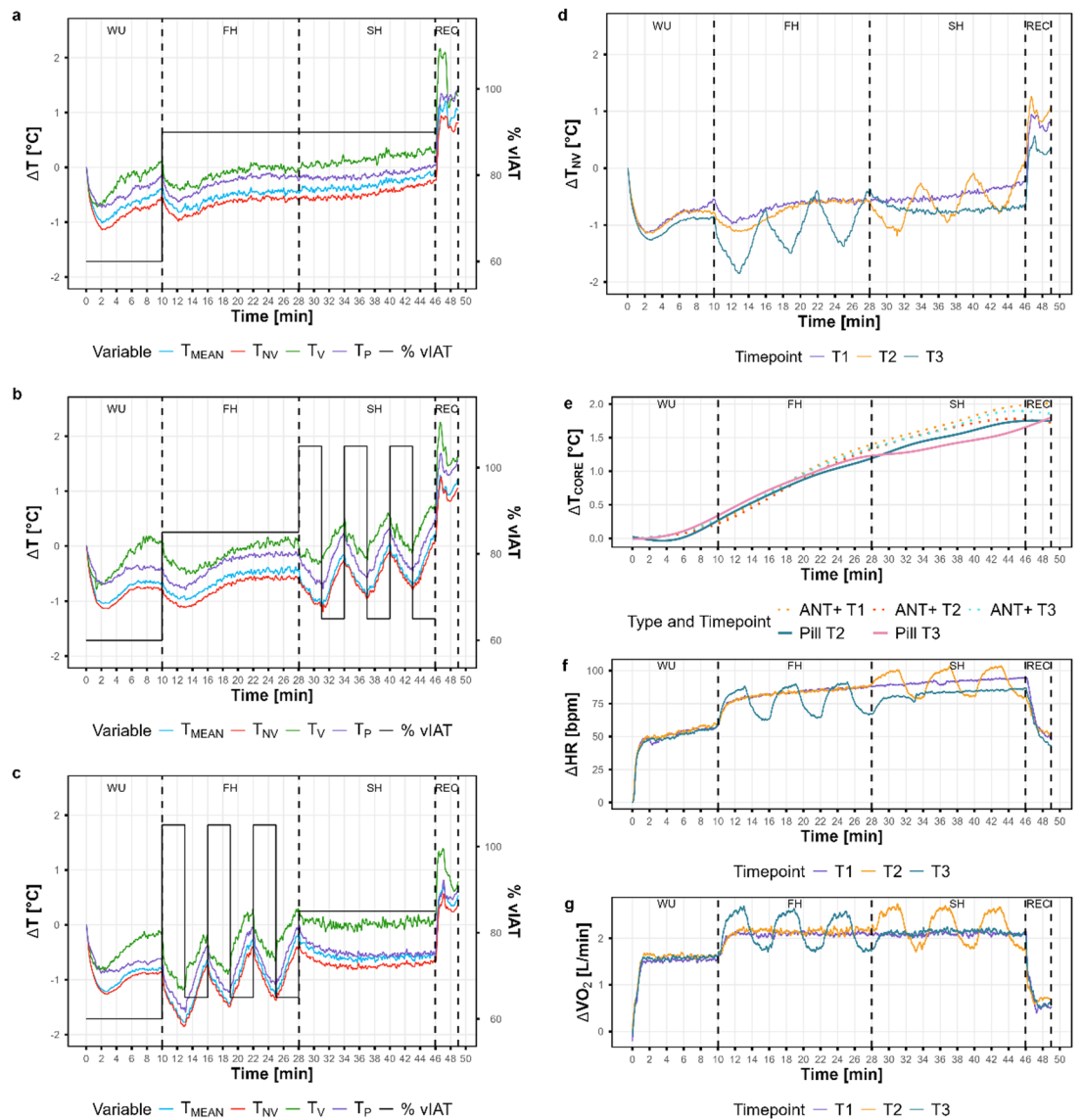
### Response of $T_{SK}$ to varying external load in comparison to established physiological markers

Overall, all  $T_{SK}$  metrics followed the variations in external load during the three running sessions, but systematic differences were observed between the metrics (Fig. 2a–c).

The  $T_{SK}$  metrics maintained a consistent order ( $T_V > T_P > T_{MEAN} > T_{NV}$ ) throughout all running sessions. During the warm-up, all  $T_{SK}$  metrics decreased immediately within the first 2–3 min, with  $T_{NV}$  showing the greatest reduction, ranging from  $-1.13$  to  $-1.27$  °C. From approximately 3 min after exercise onset,  $T_{SK}$  increased for all metrics, with  $T_V$  showing the steepest increase and reaching approximately baseline levels by the end of the warm-up period. In contrast,  $T_{NV}$  remained  $-0.55$  to  $-0.86$  °C below baseline values at the end of the warm-up phase.  $T_{SK}$  variations during the continuous phase were comparable between T1 and T2. When the continuous load was applied during the second half of the session,  $T_{NV}$  further increased in T1, whereas a decrease was observed in T3, in which the intermittent phase preceded the continuous load. During the intermittent phase, a more pronounced variation in  $T_{NV}$  was observed in T3, where participants had only a 10-min warm-up beforehand, compared to T2, where participants had already run for 28 min before the first interval ( $-0.45$  °C,  $-0.08$  °C and  $-0.26$  °C during the 1st, 2nd and 3rd interval, respectively). In line with the different load protocols, during the second half of the running session,  $T_{NV}$  showed the greatest decrease during continuous load in T3 ( $-0.66$  °C). Conversely,  $T_{NV}$  slightly increased during intermittent load in T2 ( $+0.13$  °C). All  $T_{SK}$  metrics increased immediately after exercise termination, with the highest increase after T1, ranging from 1.18 to 1.81 °C for  $T_{NV}$  and  $T_V$ , respectively. The mean temperature increases during this period for all four

<i>n</i>	Sex	Age [y]	Body height [cm]	Body mass [kg]	BMI [kg/m <sup>2</sup> ]	Body fat [%]	Lean mass [kg]	BSA [m <sup>2</sup> ]	PA-R [1–10]	vIAT [km/h]	$VO_{2max}$ [mL/min/kg]
1	m	25	178	76.3	24.1	15.1	60.8	1.94	10	16.5	65.9
2	f	27	179	68.4	21.3	15.8	53.9	1.84	10	13.3	47.5
3	m	27	178	67.6	21.3	11.6	56	1.83	8	12.2	48.8
4	m	29	181	83.6	24.1	10.2	66.6	2.05	7	12.2	59.2
5	f	22	167	57.1	20.5	20.3	42.5	1.63	9	12.5	50
6	m	22	170	70.4	24.3	18.1	54	1.82	9	13.3	57.9
7	f	23	160	60.5	23.6	21.3	44.5	1.64	10	10.7	45.9
8	m	23	177	72.6	23.1	14.1	58.4	1.89	6	11.2	56.9
9	f	26	166	58.9	21.4	25.3	41.1	1.65	7	9.9	41
10	f	24	175	65.2	21.3	13.6	52.8	1.78	7	10.4	41.8
11	f	24	177	72.2	22.8	18.1	54.7	1.88	9	12.5	49.2
Mean ± SD		24.7 ± 2.3	173.5 ± 6.7	68.4 ± 7.9	22.5 ± 1.4	16.7 ± 4.5	53.2 ± 7.8	1.8 ± 0.1	8.4 ± 1.4	12.2 ± 1.8	51.3 ± 7.8

**Table 1.** Characteristics of the participants. \*BMI = body mass index; BSA = body surface area; PA-R = physical activity rating; vIAT = velocity at individual anaerobic threshold;  $VO_{2max}$  = relative maximum oxygen uptake.



**Fig. 2.** Synchronized time series data for all three running sessions, averaged over the study cohort. X-axis represents exercise time and different load sections: warm-up (WU), first half (FH), second half (SH), recovery (REC) are separated by vertical dashed lines. The mean temperature ( $T_{MEAN}$ ), non-vessel temperature ( $T_{NV}$ ), vein temperature ( $T_V$ ) and perforator vessel temperature ( $T_P$ ) are presented for the different running sessions consisting of continuous running (T1) (a), continuous-intermittent running (T2) (b) and intermittent-continuous running (T3) (c) with the external load protocol (oriented on the running velocity at the individual anaerobic threshold [%vIAT]) as second Y-axis. The variations of  $T_{NV}$  (d), core temperature ( $T_{CORE}$ ) measured with an ANT+ sensor and telemetric pill (e), heart rate (HR) (f) and oxygen consumption ( $VO_2$ ) (g) are presented in a single panel for all three running sessions. Excluded tests due to measurement errors:  $T_{CORE}$ -ANT +  $n=1$ ,  $T_{CORE}$ -PILL  $n=1$ , HR  $n=5$ ,  $VO_2 n=1$ .

$T_{SK}$  metrics were comparable between T2 and T3 ( $1.2 \pm 0.19$  and  $1.21 \pm 0.06$  °C, respectively), whereas T1 showed a greater increase of  $+1.43 \pm 0.27$  °C. When comparing  $T_{NV}$  variations with other physiological parameters— $T_{CORE}$ , HR and  $VO_2$ —HR shows the strongest inverse association (Fig. 2e–g). The most pronounced reduction in HR during the recovery phase occurred after T1 (–48 bpm), corresponding to the greatest increase in  $T_{NV}$  after T1. In contrast, the smallest decrease in HR (–29 bpm) was observed following T2.  $T_{CORE}$  increased continuously throughout all running sessions to a similar extent in T1, T2, and T3.  $VO_2$  varied in line with the external load, with a similar magnitude of variation in all three running sessions, regardless of the prior acute load.

No differences between running sessions after each running phase (WU, FH, SH, REC) were found for  $T_{CORE}$  and  $T_{NV}$ ; however,  $T_{NV}$  data showed tendencies to differ between the tests at the end of the exercise, which are attenuated by large confidence intervals (CI; Supplementary Fig. S5). Only RPE showed a significant interaction effect between exercise session and running phase ( $p < 0.001$ ;  $\eta^2_p: 0.28$ ). Significant differences between running

sessions were observed for RPE after the first half of running in T3 compared with T1 ( $\Delta = -3.1$ ,  $SE=0.711$ ,  $p=0.002$ ) and T2 ( $\Delta = -3.5$ ,  $SE=0.711$ ,  $p<0.001$ ). At the end of the exercise sessions, RPE was significantly lower at T2 than at T1 ( $\Delta = -3$ ,  $SE=0.711$ ,  $p=0.004$ ).

Repeated-measurement correlation showed consistently strong negative correlations for all individuals during the intermittent phase between  $T_p$  and both HR (T2:  $r_{rm} = -0.7$ , 95% CI  $[-0.77, -0.6]$ ; T3:  $r_{rm} = -0.63$ , 95% CI  $[-0.71, -0.53]$ ) and  $VO_2$  (T2:  $r_{rm} = -0.72$ , 95% CI  $[-0.78, -0.64]$ ; T3:  $r_{rm} = -0.63$ , 95% CI  $[-0.7, -0.52]$ ). Furthermore, during the recovery period there were strong negative correlations in all three running sessions with  $T_p$ : T1 (HR:  $r_{rm} = -0.8$ , 95% CI  $[-0.89, -0.65]$ ;  $VO_2$ :  $r_{rm} = -0.9$ , 95% CI  $[-0.94, -0.84]$ ), T2 (HR:  $r_{rm} = -0.61$ , 95% CI  $[-0.78, -0.34]$ ;  $VO_2$ :  $r_{rm} = -0.86$ , 95% CI  $[-0.92, -0.78]$ ) and T3 (HR:  $r_{rm} = -0.8$ , 95% CI  $[-0.9, -0.63]$ ;  $VO_2$ :  $r_{rm} = -0.9$ , 95% CI  $[-0.94, -0.83]$ ). All correlations were significant at  $p<0.001$ . During the warm-up, there were consistently strong positive correlations between  $T_{CORE}$  and the entropy of  $T_p$  for all three running sessions, measured with both pill (T2:  $r_{rm} = 0.56$ , 95% CI  $[0.38, 0.7]$ ; T3:  $r_{rm} = 0.68$ , 95% CI  $[0.54, 0.78]$ ) and external ANT+ sensor (T1:  $r_{rm} = 0.72$ , 95% CI  $[0.59, 0.81]$ ; T2:  $r_{rm} = 0.59$ , 95% CI  $[0.43, 0.72]$ ; T3:  $r = 0.83$ , 95% CI  $[0.75, 0.89]$ ). All correlations were significant at  $p<0.001$ .

Individual differences were observed when comparing  $T_{CORE}$  with  $T_{NV}$  during the warm-up. The two participants with the highest vIAT exhibited consistent negative correlations regardless of the  $T_{CORE}$  measurement method. For these participants, correlation coefficients between  $T_{NV}$  and  $T_{CORE}$  pill ranged from  $r = -0.75$  to  $-0.98$  across T2 and T3 (all  $p<0.05$ ), while for  $T_{CORE}$ -ANT+, two out of the three running sessions showed statistically significant correlations, with  $r$  ranging from  $-0.7$  to  $-0.87$  (all  $p<0.05$ ). The participant with the highest vIAT (16.5 km/h) also showed strong negative correlations during the first continuous running section in T1 ( $r = -0.71$  to  $-0.79$ ,  $p<0.001$ ) and T2 ( $r = -0.89$  to  $-0.91$ ,  $p<0.001$ ) measured with both pill and ANT+. In contrast, strong positive correlations between  $T_{CORE}$ -ANT+ and  $T_{NV}$  during the warm-up were found for 6 participants at T1 ( $r = 0.83$ – $0.98$ ,  $p<0.01$ ), 7 participants at T2 ( $r = 0.72$ – $0.98$ ,  $p<0.05$ ), and at T3 ( $r = 0.81$ – $0.98$ ,  $p<0.01$ ). During the first continuous running phase, significant strong positive correlations were observed in 6 participants at T1 ( $r = 0.72$ – $0.99$ ,  $p<0.01$ ) and 9 participants at T2 ( $r = 0.71$ – $0.99$ ,  $p<0.01$ ), including one of the two individuals who showed negative correlations during the warm-up.

### Environmental and physiological factors affecting $T_{SK}$ variability

Further exploratory analyses revealed several significant correlations between  $T_{SK}$  variations and physiological factors. Figure 3 presents the four most consistent correlations across the running sessions.

The strongest correlation was found between  $\Delta T_p$  and sweat loss during the FH of the running sessions, with very high consistency across running sessions ( $r = -0.85$  to  $-0.91$ ,  $p \leq 0.001$ ). In the same running phase, a strong negative correlation between  $\Delta T_{NV}$  and the FEV<sub>1</sub> was detected across all three running sessions ( $r = -0.82$  to  $-0.89$ ,  $p \leq 0.002$ ).  $\Delta T_{NV}$  during the Pre-FH phase decreased more with higher vIAT in all three running sessions ( $r = -0.74$  to  $-0.88$ ,  $p \leq 0.009$ ). A strong negative correlation was also found between the  $\Delta T_p$  in the Pre-FH phase and MHP at exercise termination ( $r = -0.63$  to  $-0.79$ ,  $p \leq 0.038$ ). There were strong positive correlations between  $T_{CORE}$ -pill at exercise termination and both  $\Delta T_p$  from FH to recovery ( $r = 0.70$ – $0.78$ ,  $p \leq 0.017$ ) and  $\Delta T_p$  during recovery ( $r = 0.62$ – $0.64$ ,  $p \leq 0.045$ ).  $T_{CORE}$ -ANT+ at exercise termination showed similar correlations with  $\Delta T_p$  from warm-up to recovery ( $\Delta TP$  WU-REC) at T3 ( $r = 0.85$ ,  $p = 0.002$ ) and a near-significant correlation at T2 ( $r = 0.56$ ,  $p = 0.07$ ), but no association at T1 ( $r = 0.04$ ,  $p = 0.906$ ).  $VO_{2max}$  showed a strong negative correlation with  $\Delta T_{MEAN}$  during the warm-up at T1 and T2 ( $r = -0.61$  to  $-0.76$ ,  $p = 0.047$ – $0.014$ ), whereas the correlation at T3 was not significant ( $r = -0.51$ ,  $p = 0.107$ ). Additionally,  $VO_{2max}$  correlated significantly with  $\Delta T_p$  during the first half of exercise at T1 ( $r = -0.71$ ,  $p = 0.014$ ) and T3 ( $r = -0.71$ ,  $p = 0.014$ ), but not at T2 ( $r = -0.45$ ,  $p = 0.169$ ). Environmental temperature was positively correlated with  $\Delta T_p$  Pre-REC, reaching statistical significance only at T1 ( $r = 0.74$ ,  $p = 0.009$ ; T2-T3:  $r = 0.35$ – $0.58$ ,  $p = 0.064$ – $0.317$ ). In contrast, environmental humidity showed no consistent association in any exercise phase.

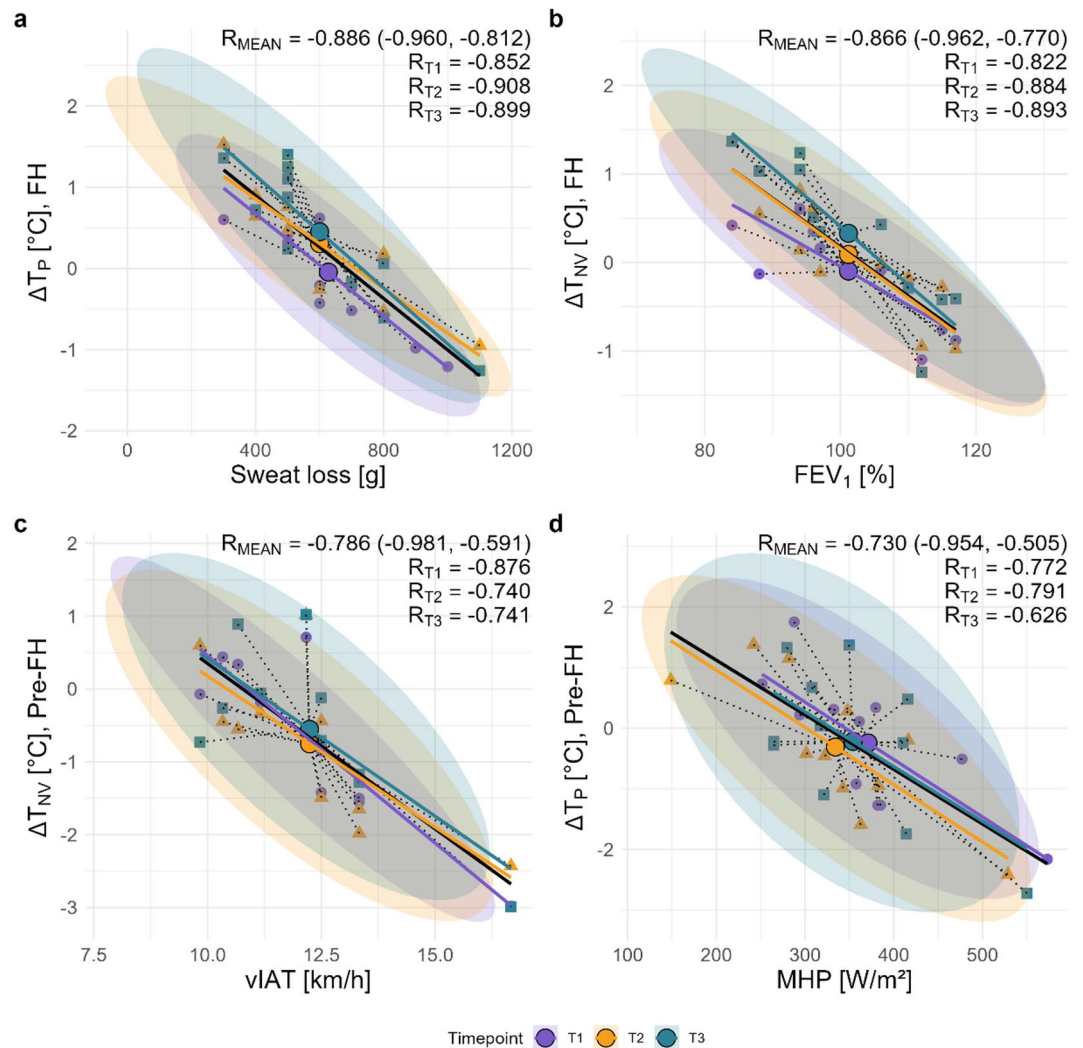
Comparing participants with high versus low vIAT revealed no significant differences in  $T_{CORE}$  and HR, but significantly higher  $T_{CORE}$ - $T_{NV}$  gradients at the end of the first half and at exercise termination were observed in all running sessions, measured with an ANT+ sensor ( $p<0.05$ , Fig. 4). In contrast, the  $T_{CORE}$ - $T_{NV}$  gradient did not differ significantly between individuals with high and low  $VO_{2max}$  during exercise.

### $T_{SK}$ reproducibility in repeated running sessions

Overall, good to excellent reproducibility was found for the recovery section [ICC(3,1): 0.91 (95% CI 0.9–0.93)], with 8/11 participants showing excellent ICC coefficients ( $\geq 0.9$ ); for the intermittent sections T2-T3 [ICC(3,1): 0.81 (95% CI 0.79–0.84)], with 5/11 participants showing excellent ICC coefficients; and for the warm-up section [ICC(3,1): 0.76 (95% CI 0.69–0.83)], with 6/11 participants showing excellent ICC coefficients. Moderate ICCs were found for the continuous section T1-T2 [ICC(3,1): 0.61 (95% CI 0.57–0.66)]. Lower ICCs were observed for the continuous sections T1-T3 [ICC(3,1): 0.14 (95% CI 0.07–0.21)] and T2-T3 [ICC(3,1): 0.14 (95% CI 0.07–0.21)], because  $T_{NV}$  values during T3 decreased in the continuous section but increased in the corresponding comparison sections. In general, ICC coefficients were higher when analyzing consistency [ICC(3,1): 0.56 (95% CI 0.44–0.68)] compared with agreement [ICC(2,1): 0.52 (95% CI 0.42–0.62)] across all running sections (Fig. 5a and b). Individual time series with ICC(3,1) and ICC(2,1) during every running section can be found as Supplementary Figs. S1 and S2. The inter-individual comparison showed good consistency, with ICC(3,1)  $\geq 0.75$  for all three running sessions (Fig. 5c).

### Discussion

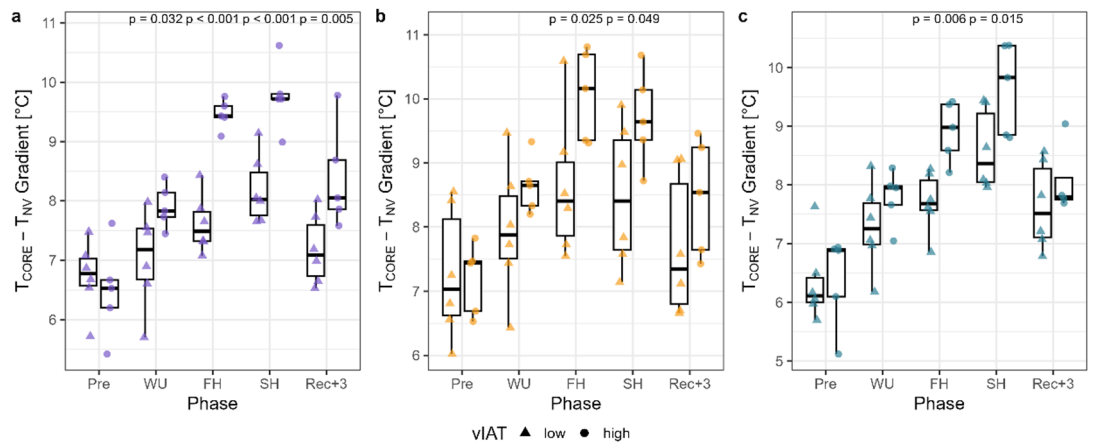
To the best of our knowledge, this is the first study to continuously examine different variables of skin surface radiation temperature, automatically detected using deep learning-based analysis of thermograms during intermittent and continuous running in endurance-trained individuals. We compared synchronized time series



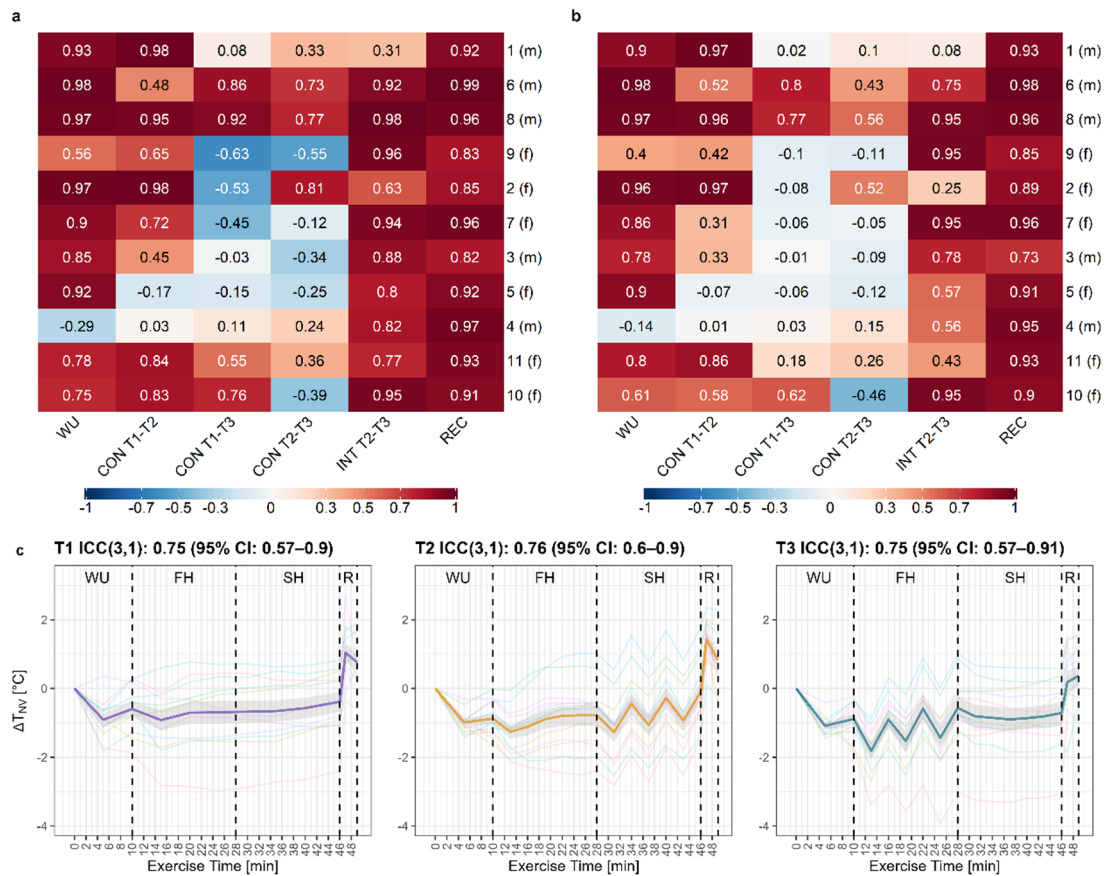
**Fig. 3.** Associations between differences in non-vessel ( $\Delta T_{NV}$ ) and perforator vessel ( $\Delta T_p$ ) temperature with potential physiological influencing parameters within certain sections of the running protocols. **(a)**  $\Delta T_p$  during the first exercise half (FH) vs. sweat loss [g]. **(b)**  $\Delta T_{NV}$  during the first exercise half (FH) vs. forced expiratory flow during the first second (FEV<sub>1</sub>) [%]. **(c)**  $\Delta T_{NV}$  from baseline to the end of the first exercise half (Pre-FH) vs. running velocity at the individual anaerobic threshold (vIAT) [km/h]. **(d)**  $\Delta T_p$  from baseline to the end of the first exercise half (Pre-FH) vs. metabolic heat production (MHP) [W/m<sup>2</sup>] at end of the running session. Data points were color-coded by timepoint (T1, T2, T3), and 95% confidence ellipses were drawn to indicate the distribution of observations within each timepoint. Centroids for each timepoint were calculated and displayed, and the mean correlation coefficients ( $R_{MEAN}$ ) with 95% confidence intervals and correlation coefficients for each individual timepoint ( $R_{T1}$ ,  $R_{T2}$ ,  $R_{T3}$ ) were annotated in the top right corner of each plot.

data of four  $T_{SK}$  metrics ( $T_{MEAN}$ ,  $T_{NV}$ ,  $T_V$ ,  $T_p$ ),  $T_{CORE}$ , RPE, and cardiorespiratory response (HR,  $VO_2$ ) during three repeated running sessions with a similar average external load but differing in acute load variation. Based on our analysis, all  $T_{SK}$ -derived metrics varied in line with changes in external load. The entropy and mean of  $T_p$  showed stronger correlations with HR,  $VO_2$  and  $T_{CORE}$  than other  $T_{SK}$  metrics. Furthermore,  $T_{SK}$  measurements during exercise were reproducible at similar external load regardless of day, time of day, or prior exercise session under standardized environmental conditions, and inter-individual differences in  $T_{SK}$  variations were associated with submaximal exercise capacity.

When analyzing  $T_{SK}$  in relation to external load, the relative variation of all  $T_{SK}$  metrics responded immediately and inversely to changes in external load during all three tests, with consistent but distinct absolute values between the  $T_{SK}$  metrics ( $T_V > T_p > T_{MEAN} > T_{NV}$ ). The temperature decrease after exercise onset and the subsequent increase after exercise termination observed across all  $T_{SK}$  metrics are consistent with mean temperature responses reported in previous studies<sup>8,53</sup>. Additionally, the spatial differentiation aligns with findings by Arfaoui et al.<sup>54</sup>, who showed that perforator vessels in the gastrocnemius muscle area exhibited higher temperatures than the surrounding ROI during graded cycling exercise. Although all  $T_{SK}$  metrics appeared to follow a similar global trend, the analysis revealed differences in their correlation with cardiopulmonary variables. The entropy and mean of  $T_p$  showed a strong correlation to cardiorespiratory parameters (e.g. HR,



**Fig. 4.** Gradients of core temperature ( $T_{CORE}$ ) and non-vessel temperature ( $T_{NV}$ ) at baseline (Pre), after the warm-up (WU), the first exercise phase (FH), the second exercise phase (SH) and the 3 min recovery phase (REC + 3) between individuals with high and low running speed at the individual anaerobic threshold (vIAT, median split) in running session T1 (a), T2 (b) and T3 (c).  $T_{CORE}$  was measured with the ANT+ sensor. Statistical significance:  $p < 0.05$  (linear mixed effects model with Tukey's post hoc test).



**Fig. 5.** Intra- and inter-individual intra-class correlation (ICC) analysis of the non-vessel temperature ( $T_{NV}$ ). Intra-individual ICC(3,1) for consistency (a) and ICC(2,1) for agreement (b) during the warm-up (WU), continuous running (CON), intermittent running (INT) and the recovery phase (REC) for each individual as separate row. (c) Inter-individual correlation analysis with mean time series and 95% confidence intervals (shaded area) for T1, T2 and T3.  $n = 3$  participants are missing  $T_{NV}$  data during the recovery phase in T3.

VO<sub>2</sub>) during different phases of the running sessions. Similarly, Bogomilsky et al.<sup>11</sup> and Hu et al.<sup>17</sup>, reported an increase in entropy of the chest surface area that was significantly correlated with exercise duration and intensity during incremental cycling. Furthermore, Masur et al.<sup>55</sup>, reported stronger associations between the kinetics of the thermal contrast index, related to activated perforasomes and internal load parameters than with mean T<sub>SK</sub> during and following incremental cycling. The high-frequency IRT measurement in the present study revealed an inverse correlation between the entropy of T<sub>p</sub> and HR and VO<sub>2</sub> during the intermittent running. This finding suggests short-term alterations in perforator vessels tone between short-term sympathetically mediated vasoconstriction during increases in external load and local vasodilation when external load is reduced, suggesting increased peripheral heat dissipation. Together these findings emphasize the potential for further investigation into different perforasome-associated metrics in the field of endurance exercise, particularly when combined with high-frequency IRT measurements for evaluating short-term physiological alterations during exercise. In contrast to these rapid, load-sensitive T<sub>SK</sub> variations, T<sub>CORE</sub> increased continuously throughout the running sessions and reflects cumulative rather than acute thermal strain. Consequently, concurrent monitoring of T<sub>SK</sub> and T<sub>CORE</sub> enables differentiation between acute peripheral heat dissipation mechanisms and systemic heat accumulation—a distinction that is, however, highly dependent on environmental context.

The thermoneutral conditions of our experiments ensured that ambient temperature did not confound the T<sub>SK</sub> metrics, providing a controlled baseline for evaluating the exercise-induced thermoregulatory response. Recently, Martínez-Noguera et al.<sup>56</sup> observed region-dependent T<sub>SK</sub> dynamics in racewalkers, with more pronounced T<sub>SK</sub> decrease at 17 °C compared to stable or attenuated T<sub>SK</sub> responses at 28 °C, indicating that pre-existing cutaneous vasodilation under heat stress alters the response of IRT-derived metrics during exercise. Likewise, Vainer<sup>57</sup> reported structurally comparable perforasome patterns after running and sauna exposure, indicating that both metabolically and externally induced thermal stress activate overlapping superficial vascular networks detectable by IRT. Hence, the thermoneutral conditions of the present study likely facilitated more pronounced and less confounded exercise induced T<sub>SK</sub> responses compared to what would be expected under environmental heat stress, where pre-existing cutaneous vasodilation may attenuate the exercise-induced thermoregulatory response. Notably, exercise modality introduces an additional layer of specificity: Hillen et al.<sup>3</sup> observed that strength training produces superficial vein prominence, whereas endurance exercise and sauna exposure activate perforator vessel patterns—distinct vascular signatures that our deep learning approach is well positioned to differentiate. Together, these findings suggest that the appropriate T<sub>SK</sub> metric and its interpretations should be selected with awareness of both exercise mode and environmental conditions, and that the parameters developed here provide a flexible, non-invasive framework for capturing these modality- and context-specific thermoregulatory responses. In the context of these environmental conditions, considerable inter-individual differences in T<sub>SK</sub> responses were identified that warrant closer examination.

During the warm-up, we observed a positive correlation between the entropy of T<sub>p</sub> and T<sub>CORE</sub> across all three running sessions. Interestingly, individual differences emerged when correlating T<sub>NV</sub> with T<sub>CORE</sub> during the warm-up and the first continuous running section. Individuals with higher vIAT showed inverse correlations between T<sub>NV</sub> and T<sub>CORE</sub>, whereas strong positive correlation were found for most individuals with lower submaximal exercise capacity. This is consistent with the much higher standard deviations of T<sub>NV</sub> during the first primary running phase compared with the second, suggesting different thermoregulatory mechanisms during the early phase of exercise (within the first 28 min), possibly due to a more efficient vasomotor response. In contrast, T<sub>CORE</sub> exhibited the least inter-individual variability, reflecting a more homeostatic regulation through individual peripheral thermoregulatory mechanisms. To quantify which physiological or environmental characteristics most strongly influence these inter-individual differences, we examined a broader range of potential determinants of T<sub>SK</sub> variation. Among these, sweat loss, FEV<sub>1</sub>, vIAT and MHP showed the highest negative correlation coefficients, and except FEV<sub>1</sub>, all are closely related to the body's response to aerobic effort. While both vIAT and VO<sub>2max</sub> reflect individual endurance performance, vIAT showed a stronger association with T<sub>NV</sub> ( $r = -0.749$ , 95% CI [-0.55, -0.87]) than VO<sub>2max</sub> ( $r = -0.503$ , 95% CI [-0.19, -0.72]), consistent with the grouped differences observed above. This is in line with evidence that it is not cardiorespiratory fitness (e.g. VO<sub>2max</sub>) per se that is associated with better heat dissipation, but rather exercise training itself, or more specifically, the repeated thermal stress experienced during training sessions<sup>58</sup>. Accordingly, Samoljanic et al.<sup>59</sup> reported that running economy, independent of aerobic fitness, alters thermoregulatory responses, with individuals with low running economy showing higher increases in T<sub>CORE</sub>. Furthermore, vIAT is inherently more running-specific than VO<sub>2max</sub>—individuals with higher vIAT may spend more time training by running than by cycling or swimming, as these latter activities facilitate greater heat loss through conduction and convection, potentially further enhancing their heat dissipation capacity during running specifically. In previous studies, no differences in  $\Delta T_{CORE}$  were observed between trained and untrained individuals when exercising at a fixed metabolic heat production rate<sup>60</sup> but skin blood flow has been shown to be elevated in trained compared with untrained at the same relative external load<sup>61</sup>. The finding of higher T<sub>CORE</sub> - T<sub>NV</sub> gradients in individuals with higher submaximal exercise capacity across all three running sessions in the present study highlights the importance of T<sub>SK</sub> measurements for understanding variability in peripheral short-term mechanisms that ultimately lead to similar increases in T<sub>CORE</sub>. This finding is in line with Périard et al.<sup>30</sup> who reported higher core-to-skin gradients in trained versus untrained individuals. Narrowing these gradients has been shown to increase cardiovascular strain (e.g. increased HR) during moderate intensity cycling<sup>31</sup>. In our cohort, an increased HR response was also observed in the group with lower vIAT, although this did not reach statistical significance. Thus, individuals with higher submaximal exercise capacity appear to exhibit better-developed peripheral thermoregulatory mechanisms during submaximal running exercise than individuals with lower submaximal exercise capacity and therefore may seem to experience reduced cardiovascular strain. While these inter-individual differences are physiologically meaningful, the practical value of continuous T<sub>SK</sub> monitoring as a tool ultimately depends on its reproducibility across sessions and individuals.

The relative  $T_{NV}$  variations during running were reproducible both intra- and inter-individual, regardless of daytime or immediate prior load on the same day, for running sections of identical (warm-up, recovery) or substantial variations in external load (intermittent). However, the reproducibility was weak when comparing continuous load sections of the primary exercise phase. Although the external workload was comparable across conditions, the load was increased immediately prior to the continuous phase in T3, which likely contributed to a reduction in  $T_{NV}$  due to acute short-term vasoconstriction. In contrast, no preceding load increase occurred in T1, and in T2 the continuous phase was positioned earlier within the primary exercise segment. These differences in sequencing and prior loading complicate direct comparisons across conditions. In the majority of previous studies, the reproducibility of IRT measurements has been reported only for resting conditions, with ICCs ranging between 0.4 and 0.9<sup>62,63</sup>, without specifying a particular ICC model. One of the highest values (ICC=0.99) was obtained using computer-aided analysis of thermograms<sup>64</sup>, highlighting the value of automation. During running exercise under hot environmental conditions, James et al.<sup>65</sup> reported lower ICC values of 0.56 for  $T_{SK}$  measurements with IRT in comparison to a telemetry thermistor system (ICC=0.84) and hard-wired thermistor system (ICC=0.62). However, that study used a thermal camera model with a substantially lower resolution (160 × 120 pixels), and all measurements were conducted with a hand-held device, factors that may also have contributed to the lower reproducibility compared with the present study. As shown by Machado et al.<sup>66</sup>, the reproducibility of  $T_{SK}$  measurements can vary between different thermal imaging camera types. Collectively, these points highlight the necessity for further standardization in the field of thermal imaging in exercise physiology, including IR camera specifications, thermogram calibration, measurement technique, and analysis strategies.

Several influential factors affected the reproducibility of the  $T_{SK}$  measurements. From an inter-individual perspective, cohorts with different physiological profiles (e.g., endurance vs. sprint athletes) have been shown to have different  $T_{SK}$  variations during running exercise<sup>53</sup>. Therefore, higher inter-individual ICC coefficients might be achieved by investigating a more physiological homogeneous sample (e.g., reduced variability in  $VO_{2peak}$  or vIAT). Interestingly, running phases characterized by greater variability in  $T_{NV}$  (warm-up, intermittent, recovery) showed higher ICC than running phases with less variability in  $T_{NV}$  (i.e., continuous running during the primary exercise phase). This decline in ICC in low-variability segments is a well-known statistical phenomenon: when a variable lacks sufficient variation, the ICC becomes more strongly influenced by random measurement noise<sup>67</sup>. Additionally, three different time series were included for each participant in the analysis of the warm-up and recovery phases, compared with two time series during the continuous running phases, making the ICC estimates for continuous running more sensitive to random error due to fewer data points.

Future studies should use high-frequency  $T_{SK}$  measurements, as static end-point values fail to capture individual short-term thermal responses. This is also supported by the systematic review of Rojas-Valverde et al.<sup>2</sup>, which documented the presence of heterogeneous inter-individual  $T_{SK}$  variations. In addition to comparisons of group means, the robust findings on individual differences support individual-specific analyses. Research in thermal imaging requires further standardisation. Addressing methodological discrepancies will enhance study reliability and comparability and advances research in exercise thermoregulation and performance monitoring. Future research could implement reinforcement or unsupervised learning techniques in the DNN-driven processing pipeline to further improve automated vessel detection and the differentiation of different  $T_{SK}$  metrics. To generalise our results, the application of DNN-assisted IRT should be replicated in different contexts, including healthy and pathological individuals (e.g., vascular diseases). The subsequent progression in AI-driven analysis will facilitate the automated identification of diseases through artificial classification, while ensuring the preservation of explainability for users.

Despite the strengths of this study, several limitations must be considered when interpreting the findings. The main findings may be generalizable only to young, healthy, and endurance-trained individuals and should be replicated in larger samples with more diverse characteristics. This study was conducted in thermoneutral conditions to focus on exercise-induced thermoregulatory responses. While this approach is consistent with most studies examining skin temperature during exercise, it limits the generalizability of our findings to exercise performed under heat stress or in the cold. Automatic ROI detection during high-velocity running would benefit from improved IR detector capabilities, such as shorter integration times or different shutter types. These improvements would reduce motion blur, missing data points, and noise, leading to enhanced pattern recognition. The image processing pipeline automatically provides objective, reproducible data on predefined features and parameters extracted from thermal images. Nevertheless, investigators must still verify the data output, which requires expert knowledge.

## Conclusion

For the first time, deep learning-assisted IRT provided valuable insights into the interplay and intercorrelation between synchronized skin temperature, core body temperature, perceived exertion, and cardiopulmonary responses during running exercise. The automated thermogram analysis showed consistent intra- and inter-individual  $T_{SK}$  variations during repeated running sessions. The robust associations between skin temperature and thermal entropy of the perforator vessels, and body core temperature and sweat loss, highlights the potential for further investigation of skin temperature and its derived metrics. Finally, inter-individual variability in thermoregulatory responses to running exercise appears to be more closely related to submaximal, running-specific performance capacity than to maximal aerobic capacity.

## Data availability

The data presented in this study are not publicly available but are available on reasonable request from the corresponding author.

Received: 21 November 2025; Accepted: 9 March 2026

Published online: 28 March 2026

## References

- Hillen, B. et al. Acute physiological responses to a pyramidal exercise protocol and the associations with skin temperature variation in different body areas. *J. Therm. Biol.* **115**, 103605 (2023).
- Rojas-Valverde, D. et al. Short-term skin temperature responses to endurance exercise: A systematic review of methods and future challenges in the use of infrared thermography. *Life* **11**, 1–27 (2021).
- Hillen, B., Pfirrmann, D., Nägele, M. & Simon, P. Infrared thermography in exercise physiology: The dawning of exercise radiomics. *Sports Med.* **50**, 263–282 (2020).
- Tsoutsoubi, L., Ioannou, L. G., Cheung, S. S. & Flouris, A. D. Influence of core temperature changes during whole-body warming and cooling on cutaneous vascular reactivity. *J. Hum. Perform. Extreme Environ.* **18**, (2023).
- Masur, L., Brand, F. & Düking, P. Response of infrared thermography related parameters to (non-)sport specific exercise and relationship with internal load parameters in individual and team sport athletes—a systematic review. *Front. Sports Act. Living.* **6**, 1479608 (2024).
- Drzazga, Z., Binek, M., Pokora, I. & Sadowska-Krepa, E. A preliminary study on infrared thermal imaging of cross-country skiers and swimmers subjected to endurance exercise. *J. Therm. Anal. Calorim.* **134**, 701–710 (2018).
- Cabizosu, A., Marin-Pagan, C., Alcaraz, P. E. & Martínez-Noguera, F. J. Infrared thermography sensor in the analysis of acute metabolic stress response during race walking competition. *Biosensors* **14**, 478 (2024).
- Tanda, G. The use of infrared thermography to detect the skin temperature response to physical activity. *J. Phys. Conf. Ser.* **655**, 012062 (2015).
- Jones, T. W., Shillabeer, B. C. & Cardinale, M. Skin temperature, training load, and subjective muscle soreness in junior endurance athletes: A case study. *Int. J. Sports Physiol. Perform.* **15**, 1349–1352 (2020).
- Aylwin, P. E. et al. The use of infrared thermography for the dynamic measurement of skin temperature of moving athletes during competition; methodological issues. *Physiol. Meas.* **42**, 084004 (2021).
- Bogomilsky, S., Hoffer, O., Shalmon, G. & Scheinowitz, M. Preliminary study of thermal density distribution and entropy analysis during cycling exercise stress test using infrared thermography. *Sci Rep.* **12**, (2022).
- Hunold, S., Mietzsch, E. & Werner, J. Thermographic studies on patterns of skin temperature after exercise. *Eur. J. Appl. Physiol.* **65**, 550–554 (1992).
- Belinchón-deMiguel, P., Ramos-Campo, D. J. & Clemente-Suárez, V. J. Physiological, biomechanical, and thermographic responses in male athletes during an ultra-endurance race. *Appl. Sci.* **14**, 6511 (2024).
- Silva, A. G. et al. Effect of whole-, upper-, and lower-body high-intensity rowing exercise on skin temperature measured by thermography. *Res. Q. Exerc. Sport.* **94**, 226–236 (2023).
- Kapoor, M., Vasdev, V., Singh, R. K., Jaipurkar, R. & Sikri, G. Relationship between aerobic fitness and lower limb skin temperature during cycling exercise testing among well-trained athletes and nonathletes: A cross-sectional study. *Med. J. Armed Forces India.* **79**, S165–S174 (2023).
- Jastrzębska, A. D., Hebisz, R. & Hebisz, P. Temporal skin temperature as an indicator of cardiorespiratory fitness assessed with selected methods. *Biology* **11**, 948 (2022).
- Hu, C., Du, N., Liu, Z. & Song, Y. Can infrared thermal imaging reflect exercise load? An incremental cycling exercise study. *Bioengineering* **12**, 280 (2025).
- Novotny, J., Rybarova, S., Zacha, D., Bernacikova, M. & Ramadan, W. A. The influence of breaststroke swimming on the muscle activity of young men in thermographic imaging. *Acta Bioeng. Biomech.* **17**, 121–129 (2015).
- Morán-Navarro, R. et al. Validity of skin, oral and tympanic temperatures during exercise in the heat: Effects of wind and sweat. *Ann. Biomed. Eng.* **47**, 317–331 (2019).
- Fernandes, A. D. A. et al. Measuring skin temperature before, during and after exercise: A comparison of thermocouples and infrared thermography. *Physiol. Meas.* **35**, 189–203 (2014).
- Viegas, F. et al. The use of thermography and its control variables: A systematic review. *Rev. Bras. Med. Esporte.* **26**, 82–86 (2020).
- Perpetuini, D., Formenti, D., Cardone, D., Filippini, C. & Merla, A. Regions of interest selection and thermal imaging data analysis in sports and exercise science: A narrative review. *Physiol. Meas.* **42**, 08TR01 (2021).
- Priego Quesada, J. I., Lucas-Cuevas, A. G., Palmer, S. & Pérez-Soriano, R. Cibrián Ortiz De Anda, R. M. Definition of the thermographic regions of interest in cycling by using a factor analysis. *Infrared Phys. Technol.* **75**, 180–186 (2016).
- Costa, C. M. A. et al. Daily oscillations of skin temperature in military personnel using thermography. *J. R Army Med. Corps.* **162**, 335–342 (2016).
- Korman, P. et al. Exploring the correlation of skin temperature and body composition in athletes undergoing exhaustive physical exercise. *J. Therm. Biol.* **123**, 103918 (2024).
- Neves, E. B. et al. The influence of subcutaneous fat in the skin temperature variation rate during exercise. *Res. Biomed. Eng.* **31**, 307–312 (2015).
- Fernandes, A. A. et al. Validity of inner canthus temperature recorded by infrared thermography as a non-invasive surrogate measure for core temperature at rest, during exercise and recovery. *J. Therm. Biol.* **62**, 50–55 (2016).
- Mah, A. J., Zadeh, G., Khoshnam Tehrani, L., Askari, M. & Shadgan, B. S. Studying the accuracy of infrared thermography for measuring core body temperature. In *Biophotonics in Exercise Science, Sports Medicine, Health Monitoring Technologies, and Wearables III* (eds Shadgan, B. & Gandjbakhche, A. H.) 15 (SPIE, San Francisco, United States, 2022). <https://doi.org/10.1117/12.2608929>.
- Cuddy, J. S., Hailes, W. S. & Ruby, B. C. A reduced core to skin temperature gradient, not a critical core temperature, affects aerobic capacity in the heat. *J. Therm. Biol.* **43**, 7–12 (2014).
- Périard, J. D., Caillaud, C. & Thompson, M. W. The role of aerobic fitness and exercise intensity on endurance performance in uncompensable heat stress conditions. *Eur. J. Appl. Physiol.* **112**, 1989–1999 (2012).
- Chou, T. H., Allen, J. R., Hahn, D., Leary, B. K. & Coyle, E. F. Cardiovascular responses to exercise when increasing skin temperature with narrowing of the core-to-skin temperature gradient. *J. Appl. Physiol.* **125**, 697–705 (2018).
- Amson, H. et al. Core-to-skin temperature gradient measured by thermography predicts day-8 mortality in septic shock: A prospective observational study. *J. Crit. Care.* **60**, 294–299 (2020).
- Hillen, B., Andrés López, D., Schömer, E., Nägele, M. & Simon, P. Towards exercise radiomics: Deep neural network-based automatic analysis of thermal images captured during exercise. *IEEE J. Biomed. Health Inf.* 1–11. <https://doi.org/10.1109/JBHI.2022.3186530> (2022).
- Andrés López, D., Hillen, B., Nägele, M., Simon, P. & Schömer, E. StereoThermoLegs: Label propagation with multimodal stereo cameras for automated annotation of posterior legs during running at different velocities. *J. Therm. Anal. Calorim.* **149**, 11349–11360 (2024).
- Andrés López, D., Hillen, B., Nägele, M., Simon, P. & Schömer, E. ThermoNet: Advanced deep neural network-based thermogram processing pipeline for automatic time series analysis of specific skin areas in moving legs. *J. Therm. Anal. Calorim.* **149**, 11337–11348 (2024).

36. Hillen, B. et al. An exploratory, intra- and interindividual comparison of the deep neural network automatically measured calf surface radiation temperature during cardiopulmonary running and cycling exercise testing: A preliminary study. *J. Therm. Biol.* **113**, 103498 (2023).
37. George, J. D., Stone, W. J. & Burkett, L. N. Non-exercise VO<sub>2</sub>max estimation for physically active college students. *Med. Sci. Sports Exerc.* **29**, 415–423 (1997).
38. Borg, G. A. Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* **14**, 377–381 (1982).
39. Dickhuth, H. H. et al. Individual anaerobic threshold for evaluation of competitive athletes and patients with left ventricular dysfunction. In *Advances in Ergometry* (eds Bachl, N., Graham, T. E. & Löllgen, H.) 173–179 (Springer Berlin Heidelberg, Berlin, Heidelberg, (1991).
40. Notley, S. R., Meade, R. D. & Kenny, G. P. Time following ingestion does not influence the validity of telemetry pill measurements of core temperature during exercise-heat stress: The journal *Temperature* toolbox. *Temperature* **8**, 12–20 (2021).
41. Nishi, Y. Chapter 2 measurement of thermal balance of man. In *Studies in Environmental Science* (eds Cena, K. & Clark, J. A.) 29–39, Vol. 10 (Elsevier, 1981).
42. Mosteller, R. D. Simplified calculation of body-surface area. *N Engl. J. Med.* **317**, 1098 (1987).
43. Andrés López, D. ThermoNet: Deep neural network thermogram analysis of human calves during physical exercise. *Johannes Gutenberg-Universität Mainz*. <https://doi.org/10.25358/OPENSOURCE-10770> (2024).
44. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer Vision—ECCV 2018* (eds Ferrari, V., Hebert, M., Sminchisescu, C. & Weiss, Y.) 833–851, Vol. 11211 (Springer International Publishing, Cham, 2018).
45. Zhao, S., Wang, Y., Yang, Z. & Cai, D. *Region Mutual Information Loss for Semantic Segmentation*. Preprint at (2019). <https://doi.org/10.48550/arXiv.1910.12037>
46. Loshchilov, I. & Hutter, F. *Decoupled Weight Decay Regularization*. Preprint at (2019). <https://doi.org/10.48550/arXiv.1711.05101>
47. Oktay, O. et al. *Attention U-Net: Learning Where to Look for the Pancreas*. Preprint at (2018). <https://doi.org/10.48550/arXiv.1804.03999>
48. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *Machine Learning in Medical Imaging* (eds Wang, Q., Shi, Y., Suk, H. I. & Suzuki, K.) vol. 10541 379–387 (Springer International Publishing, Cham, (2017).
49. Zhuang, J. et al. *AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients*. Preprint at (2020). <https://doi.org/10.48550/arXiv.2010.07468>
50. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
51. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, 1988).
52. Koo, T. K. & Li, M. Y. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
53. Korman, P. et al. Response of skin temperature, blood ammonia and lactate during incremental exercise until exhaustion in elite athletes. *Sci. Rep.* **14**, 2237 (2024).
54. Arfaoui, A., Bertucci, W. M., Letellier, T. & Polidori, G. Thermoregulation during incremental exercise in masters cycling. *J. Sci. Cycl.* **3**, 33–41 (2014).
55. Masur, L., Willenbockel, H. F., Cordes, T., Sperlich, B. & Düring, P. Detection of activated perforasome patterns and the correlation with acute internal load parameters during and following incremental cycling. *Eur. J. Appl. Physiol.* <https://doi.org/10.1007/s00421-025-05943-7> (2025).
56. Martínez-Noguera, F. J., Cabizosu, A., Marín-Pagán, C. & Alcaraz, P. E. Body surface profile in ambient and hot temperatures during a rectangular test in race walker champions of the World Cup in Oman 2022. *J. Therm. Biol.* **114**, 103548 (2023).
57. Vainer, B. G. FPA-based infrared thermography as applied to the study of cutaneous perspiration and stimulated vascular response in humans. *Phys. Med. Biol.* **50**, R63–R94 (2005).
58. Ravanelli, N., Gagnon, D., Imbeault, P. & Jay, O. A retrospective analysis to determine if exercise training-induced thermoregulatory adaptations are mediated by increased fitness or heat acclimation. *Exp. Physiol.* **106**, 282–289 (2021).
59. Smoljanić, J., Morris, N. B., Derviš, S. & Jay, O. Running economy, not aerobic fitness, independently alters thermoregulatory responses during treadmill running. *J. Appl. Physiol.* **117**, 1451–1459 (2014).
60. Jay, O., Bain, A. R., Deren, T. M., Sacheli, M. & Cramer, M. N. Large differences in peak oxygen uptake do not independently alter changes in core temperature and sweating during exercise. *Am. J. Physiol. -Regul. Integr. Comp. Physiol.* **301**, R832–R841 (2011).
61. Fritzsche, R. G. & Coyle, E. F. Cutaneous blood flow during exercise is higher in endurance-trained humans. *J. Appl. Physiol.* **88**, 738–744 (2000).
62. Fernández-Cuevas, I. et al. Classification of factors influencing the use of infrared thermography in humans: A review. *Infrared Phys. Technol.* **71**, 28–55 (2015).
63. Calvo-Lobo, C. et al. Intra- and inter-session reliability and repeatability of an infrared thermography device designed for materials to measure skin temperature of the triceps surae muscle tissue of athletes. *PeerJ* **11**, e15011 (2023).
64. Fernandez Cuevas, I. et al. Reliability and reproducibility of skin temperature of overweight subjects by an infrared thermography software designed for human beings. In *Proceedings of the 12th European Congress of Thermology 130–137* (European Association of Thermology and the Austrian Society of Thermology, Porto, Portugal, 2012).
65. James, C. A., Richardson, A. J., Watt, P. W. & Maxwell, N. S. Reliability and validity of skin temperature measurement by telemetry thermistors and a thermal camera during exercise in the heat. *J. Therm. Biol.* **45**, 141–149 (2014).
66. Machado, Á. S. et al. Influence of infrared camera model and evaluator reproducibility in the assessment of skin temperature responses to physical exercise. *J. Therm. Biol.* **98**, 102913 (2021).
67. Mehta, S. et al. Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. *Stat. Med.* **37**, 2734 (2018).

## Acknowledgements

We gratefully thank all individuals who participated in the here presented experiments.

## Author contributions

Conceptualisation, B.H., P.S., E.S. and E.W.I.N.; Data collection, V.W., B.H., D.T.O and S.Z.; Data curation, V.W., B.H. and D.A.L.; Data interpretation, V.W. and B.H.; Deep neural network analysis, D.A.L.; Visualization, V.W. and D.A.L.; Writing—original draft preparation, V.W. and B.H.; Writing—Review & Editing, V.W., B.H., D.A.L., E.S., E.W.I.N., D.T.O., M.N., S.Z. and P.S.; Supervision, P.S. All authors have read and agreed to the final version of the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL. The authors gratefully announce that this work was partially funded by the Federal Ministry for Economic Affairs and Energy based on a decision of the Ger-

man Parliament (ZF4211603GR9) and partially funded by the Carl-Zeiss-Stiftung (Multi-dimensionAI project: P2022-08-010).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-44102-6>.

**Correspondence** and requests for materials should be addressed to B.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026