

JoulesEye: Energy Expenditure Estimation and Respiration Sensing From Thermal Imagery While Exercising

RISHIRAJ ADHIKARY, IIT Gandhinagar, Gandhinagar, India

MAITE SADEH, Cornell University, NY, United States

NIPUN BATRA, IIT Gandhinagar, Gandhinagar, India

MAYANK GOEL, Carnegie Mellon University, PA, United States

Smartphones and smartwatches have contributed significantly to fitness monitoring by providing real-time statistics, thanks to accurate tracking of physiological indices such as heart rate. However, the estimation of calories burned during exercise is inaccurate and cannot be used for medical diagnosis. In this work, we present *JoulesEye*, a smartphone thermal camera-based system that can accurately estimate calorie burn by monitoring respiration rate. We evaluated *JoulesEye* on 54 participants who performed high intensity cycling and running. The mean absolute percentage error (MAPE) of *JoulesEye* was 5.8%, which is significantly better than the MAPE of 37.6% observed with commercial smartwatch-based methods that only use heart rate. Finally, we show that an ultra-low-resolution thermal camera that is small enough to fit inside a watch or other wearables is sufficient for accurate calorie burn estimation. These results suggest that *JoulesEye* is a promising new method for accurate and reliable calorie burn estimation.

Additional Key Words and Phrases: respiration rate, metabolism, calorie estimation

ACM Reference Format:

Rishiraj Adhikary, Maite Sadeh, Nipun Batra, and Mayank Goel. 2023. JoulesEye: Energy Expenditure Estimation and Respiration Sensing From Thermal Imagery While Exercising. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 146 (December 2023), 29 pages. <https://doi.org/10.1145/3631422>

1 INTRODUCTION

Smartphones and watches come with various physiological sensors and functions that allow users to measure their health. Some of these sensors have become very accurate and enable a plethora of exciting health applications. For example, smartwatches can accurately estimate heart rate, which can be used for real-time, non-invasive monitoring of cardiovascular conditions. Researchers have also used different smartphone sensors to accurately diagnose lung diseases [42], jaundice [19, 49], traumatic brain injuries [50], and other conditions. However, smartphones and wrist wearables are not primarily designed for health sensing [51]. This makes them susceptible to inaccurate health diagnoses. Many sensors and functionalities on phones and watches that are marketed for their health benefits are often very inaccurate [69]. One such functionality is energy expenditure (EE). Several researchers have demonstrated that EE estimation can be underestimated or overestimated by wearables by over 40% [9, 21, 25, 59]. The study presented in this paper also corroborates these findings. Despite the development of new sensors every year, accurate energy expenditure measurement has proven elusive. There is a lack of a potentially ubiquitous approach to measuring EE, as accurate devices are often bulky and require masks. We

Authors' addresses: Rishiraj Adhikary, IIT Gandhinagar, Gandhinagar, India; Maite Sadeh, Cornell University, NY, United States; Nipun Batra, IIT Gandhinagar, Gandhinagar, India; Mayank Goel, Carnegie Mellon University, PA, United States.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2474-9567/2023/12-ART146 \$15.00

<https://doi.org/10.1145/3631422>

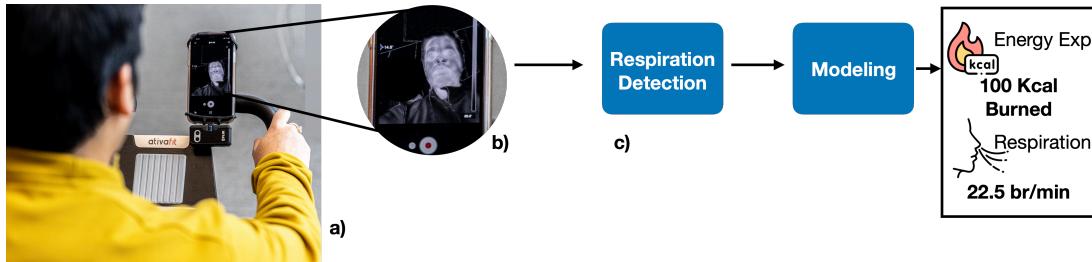


Fig. 1. *JoulesEye* estimates Energy Expenditure (EE) from respiration rate. In a) the participant is riding a cycle with thermal camera and phone fixed on the handrail. b) Shows a frame of the thermal video. c) Shows the respiration rate detection pipeline during motion combined with deep learning architecture to predict energy expenditure.

set out to identify the sensor that promises to be practical and ubiquitous enough that it could be added to a consumer device and immediately unlock accurate energy expenditure.

We present *JoulesEye* which uses thermal imaging to measure respiration to accurately estimate energy expenditure (EE) as shown in Figure 1. EE is defined as the calories burned at rest or during physical activity in a minute (kcal/min). It is one of the more frequently examined physiological outputs [69] and an accurate monitoring of EE is critical given the increasing prevalence of obesity (More than 1 billion people worldwide are obese [58]). Wearables such as a smartwatch often rely on heart rate measurement alone to estimate EE [51]. Unfortunately, studies have shown that those most dependent upon wearable devices (like people with obesity management) for EE estimation may be the ones most likely to experience inaccuracies [69].

We leverage the fact that EE depends on a multitude of factors. One such factor that is also related to respiration rate is the body composition. Body composition is defined as the proportion of fat, bone, water, and muscle in the human body. Two people with the same body weight could have very different body compositions if one has a higher amount of muscle mass and a lower amount of body fat compared to the other. Medical [43, 44, 72] and physical assessment studies [55, 56] have shown that respiration rate helps explain the change in body composition which consequently explain the change in EE. Medical grade method in estimating EE includes Double Water Test (DWT) [33], direct calorimetry and indirect calorimetry. But these methods involve bulky installation, medical supervision and participant confinement for over 24 hours [33]. The gold standard indirect calorimetry method of estimating EE requires monitoring of heart rate and the concentration of oxygen (O_2) and carbon-dioxide (CO_2) in breath.

We used a thermal camera attachment for phones to accurately estimate EE (Figure 1). Breathing causes change in temperature in the nostrils which results in variations in pixel intensity. We used classical region tracking approach like Channel and Spatial Reliability Filter [46] to retrieve the pixel intensity in the nostrils. The intensity information over time gives us the proxy of the breathing signal. We also used temperature and heart rate data to improve the results. To extract temperature, we monitored multiple points on the forehead. The forehead is an area of bony prominence where the probability of observing the change in temperature due to workout is high [34]. While prior work has estimated respiration rate using thermal images [4, 16], experiments have not been performed when the participants move exaggeratedly while exercising. Motion is also a challenge for wireless signals-based respiration monitoring as removing motion artifacts has been a longstanding challenge. Our algorithms work even when the user is cycling or running vigorously. The sensed respiration rate, temperature

and heart rate information from thermal data are fed into a deep learning model to estimate energy expenditure.

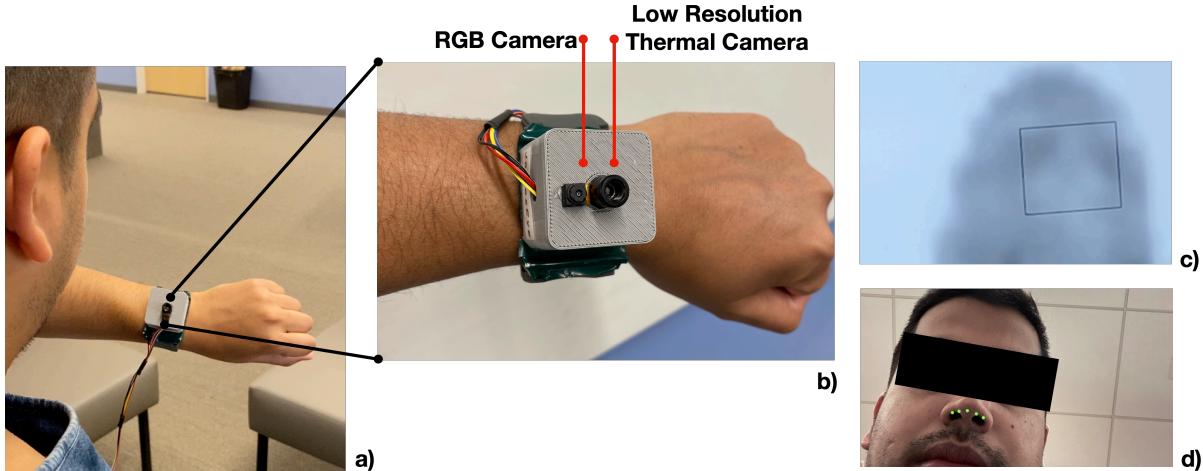


Fig. 2. a) We evaluated *JoulesEye* using low-resolution thermal camera (MLX90640). a) and b) demonstrates our vision that integrating thermal camera to smartwatch can unlock health applications. c) shows the low resolution video, where the square covers the nose and is the region of interest from where respiration signal will be extracted. The region of interest is identified using the nostril landmarks from the RGB camera as shown in d). The idea is that, a user after vigorous running or exercising can record her respiration rate for a instantaneous energy expenditure estimate. This instantaneous value in combination with heart rate will be a more accurate estimate of energy expenditure. In Section 6.6 and Section 6.5 we discuss the current challenges of using a low resolution thermal camera based smartwatch.

To evaluate our approach, we recruited 54 participants including 24 female participants who cycled and ran for 15 minutes. Our system has a Mean Absolute Percentage Error (MAPE) of 5.8% when compared with a gold standard indirect calorimeter. Given that high-resolution thermal cameras (160×120 visual resolution) are still relatively expensive, we performed an experiment to evaluate if respiration signal could be extracted using a lower resolution thermal video. We found that, our approach for estimating respiration rate works when the resolution is reduced to 32×24 pixels. We envision that low-cost, low-resolution thermal camera costing \$20 USD can be easily added to modern phone's camera cluster at the top of the screen or on a watch like shown in Figure 2 and immediately enable accurate measurement of EE.

Our work makes the following contributions:

- We created a system that can sense respiration rate and temperature from thermal video during motion and use this information to accurately estimate EE.
- We evaluate our approach on 54 participants including 24 female participants who performed cycling and running with indirect calorimetry as the ground truth. We demonstrate that respiration rate is a more accurate predictor of energy expenditure than energy expenditure calculated just using heart rate data obtained from an Apple Watch.

- We perform sensitivity analysis to show that the respiration rate estimation does not lose its accuracy when we reduce the resolution of the thermal video. Such an analysis is a step towards using ubiquitous low-resolution thermal cameras to extract energy expenditure.
- The data and code for reproducibility will be made available in the [GitHub repository](#).

2 BACKGROUND ON CALORIE ESTIMATION

Estimating a person's total energy expenditure (TEE) is critical to monitor progress towards health goals and help in preventing diseases such as diabetes, cardiovascular disease and even cancer. Any bodily movement produced by skeletal muscles results in energy expenditure [15]. We can divide total energy expenditure into three components, a) Resting Energy Expenditure (REE), b) Diet-Induced Energy Expenditure (DEE), and c) Activity-Induced Energy Expenditure (EE). REE is the energy required for growth, maintenance and normal function of the body. Significant factors that contribute to individual variation in REE include age, gender, body size, body composition, ethnicity, physical fitness level, hormonal status and various genetic and environmental influences [33]. Diet-Induced Energy Expenditure (DEE) is the energy required for food digestion, absorption, transport and metabolism, storage of nutrients, and elimination of wastes. It represents an increase in energy expenditure above the REE, which can be measured for several hours after a meal. EE is the most variable among the components of TEE. In healthy young adults, the increase in EE is a function of the training, physical activity and overall fitness.

Methods	Limitations
Double Labeled Water Test (DLW)	Requires at least 3 clinical visits in a 24-hour period.
Direct Calorimetry	Requires participant confinement for at least 24 hours.
Indirect Calorimetry	Costly device. Requires instrumenting the user with mask.
Accelerometry	Does not consider human physiology which leads to error in energy estimation.
Heart Rate	Error prone with activities of rapid intensity changes.

Table 1. There is a trade-off between accuracy and convenience of existing energy expenditure methods. Methods like DLW are very accurate but redundant. Energy Expenditure from heart rate is convenient to retrieve (for example, using smartwatches), but those estimates of energy expenditure are error prone when different activities are involved [51].

2.1 Measurement Methods of Energy Expenditure

There are several popular approaches to measuring energy expenditure during a physical activity session. Some approaches are more invasive and accurate. Typically, as the invasiveness reduces, the accuracy also reduces. Table 1 shows various methods used to measure TEE.

2.1.1 Double Labeled Water Test. The Double Labeled Water (DLW) test is a chemical-based method that is widely recognized as the gold standard for the measurement of TEE [61]. DLW test requires at least three visits to a clinic on a single day for urine sample collection at different intervals. Although the procedure itself is non-invasive, it is still tedious and time-consuming for the users to frequently measure their energy expenditure using this method.

2.1.2 Direct Calorimetry. Direct Calorimetry measures the rate of heat loss by the subject using a calorimeter¹. The subject is placed in an insulated air chamber to ensure that no heat leaks through the walls. The heat generated by the body due to metabolism changes the temperature of the space, which is then measured, usually by changes in the temperature of water flowing through the walls. This method requires subject confinement for over 24 hours and is not ideal for measuring Active Energy Expenditure (EE) as gym equipment cannot be used in a direct calorimetry chamber because the heat from gym equipment can pollute the recorded data [52].

2.1.3 Indirect Calorimetry. Indirect calorimetry is a more practical alternative to the direct method. It relies on the measurement of inspired and expired gas volume and the concentrations of O₂ and CO₂. Human energy metabolism involves the production of energy from fuel combustion in the form of carbohydrates, protein, fat or alcohol. In this process, oxygen is consumed, and carbon dioxide is produced. Indirect calorimetry is measured as a proxy of heat production or loss by measuring oxygen consumption and/or carbon dioxide production [29]. Due to its accuracy and practicality, this paper uses indirect calorimetry as the reference approach to measure energy expenditure while exercising.

2.1.4 Accelerometry. Inertial Measurement Units (IMU) can determine energy expenditure from body acceleration. Predictive equations are often used to find a relationship between body motion and energy expenditure. However, studies have shown that this approach has many limitations when translating accelerometer data to energy expenditure, primarily when used across various activities [47]. Devices that rely solely on motion data to estimate energy expenditure are unreliable because they do not consider a person's physiology such as heart and respiration rate.

2.1.5 Heart Rate. The physiological basis of using heart rate to measure energy expenditure is the estimation of consumed oxygen during aerobic metabolism. The oxygen is delivered to the working muscles and organs in blood and used to burn fat and carbohydrates to generate energy. As the body's oxygen requirement increases, so does the heart rate (HR) and stroke volume (SV), *i.e.* the volume of blood ejected from the heart to the aorta in one heartbeat. The oxygen-rich blood leaves the heart via the aorta, and the oxygen-poor blood returns to the heart via veins. It should be noted that oxygen in oxygen-rich blood is never consumed completely; thus, the blood in veins always contains some oxygen. For energy expenditure calculation, we need to thus consider the net decrease in oxygen content between oxygen-rich (arterial) and oxygen-poor (venous) blood. Oxygen consumption-based Energy Expenditure (with unit kcal/min) can be represented in one equation as:

$$EE = HR \times SV \times \delta_{av}O_2 \times \gamma$$

The heart rate is measured in beats per minute (bpm), SV is measured in litres (l), $\delta_{av}O_2$ is the ratio of litres of oxygen in the aorta to litres of oxygen in the vein or $\frac{l_a}{l_v}$. γ is the oxygen-to-energy coefficient which is $\approx 5\text{kcal/l}$ [65]. SV and $\delta_{av}O_2$ are not constant. They vary according to various parameters such as body composition [18], exercise intensity and duration.

Smartphones and Smartwatches commonly employ heart rate and accelerometry data as a surrogate for estimating Energy Expenditure (EE). In Section 3.2, we delve into the related research concerning EE estimation using consumer-grade devices.

In addition to heart rate data, respiration data helps explain change in EE because body composition and respiratory functions are correlated [43, 44, 72].

¹Example of direct calorimeter: <https://www.directindustry.com/prod/ineltec-france/product-123381-1399577.html>

2.2 Link Between Energy Expenditure and Respiration Rate:

Current energy expenditure estimation algorithms in consumer wearables do not consider individual variability, which arises because of different body compositions and varying Resting Metabolic Rate (RMR) [69]. Respiration rate can be considered a proxy measure of body composition since it is known that adiposity and respiratory functions are correlated [43, 44, 72]. Respiration rate is thus a good marker of physical effort. Also, the brain's motor cortex is a significant regulator of respiration rate [55, 57]. The same motor cortex is also the primary regulator of aerobic capacity, i.e. energy expenditure [30, 55]. Our work uses a thermal camera to estimate respiration rates. The respiration rate information is input into a deep learning model to calculate the Energy Expenditure (EE).

3 RELATED WORK

JoulesEye utilizes respiration rate data from thermal cameras to predict Energy Expenditure (EE). This section begins by exploring previous research on respiration sensing with consumer devices and wearables. We then highlight why thermal cameras outperform other methods in capturing respiration rate, especially during participant motion. Finally, we focus on earlier studies that have examined Energy Expenditure monitoring using physiological signals.

3.1 Respiration Sensing In Ubiquitous Computing

Previous research extensively utilizes the widespread availability of Inertial Measurement Units (IMUs) to detect respiration rate. IMUs have been integrated into in-ear headphones [66], glasses [32] and smartwatches [71]. However, the efficacy of detecting respiration rate using smartwatches is effective either at rest or under controlled motion, not both, due to motion-induced artifacts [11]. Research has also explored respiration rate detection through smartphones, necessitating chest strapping [6, 35]. Wireless methods employing radar signals bouncing off the human body [79] prove beneficial for detecting diseases like sleep apnea but are hampered by human motion. A comprehensive review by Ali et al. [3] covers various non-contact respiration rate detection methods, though they encounter several challenges. Vision-based systems encounter signal quality issues due to user motion, and radar systems suffer from noise due to reflections in environments with multiple objects. Now we explain in detail the reasons for not employing chest belts, RGB cameras, or earables for respiration in Energy Expenditure estimation for *JoulesEye*.

3.1.1 Respiration Sensing Using Chest belt: Force sensor embedded in the chest belts is designed to detect chest wall movements during respiration. However, existing literature [31] has highlighted a notable limitation in using chest belt data during physical activities such as cycling, where the accuracy is compromised because the oscillatory motion during cycling is at a similar frequency to the respiration rate. Furthermore, placing a sensor on the chest is not ideal for sensing respiration because minute respiration movement can be overwhelmed by the large torso movement [78]. Consistent with these findings, we also observed similar inaccuracies in chest belt readings for multiple participants engaged in running or cycling. To illustrate this point, we present a comparative example of respiration signals captured through both a thermal camera and a chest belt during motion, displayed in Figure 3. The chest belt signal is noisy, and the associated chest belt software reports a respiration rate of 5.4 breaths per minute for a 30s segment. In contrast, the thermal camera data recorded during the same period exhibits cleaner results, with a respiration rate of 22.2 breaths per minute, aligning closely with the manually determined ground truth respiration rate of 20.0 breaths per minute, obtained through visual inspection of the thermal video (explained in Section 4.5).

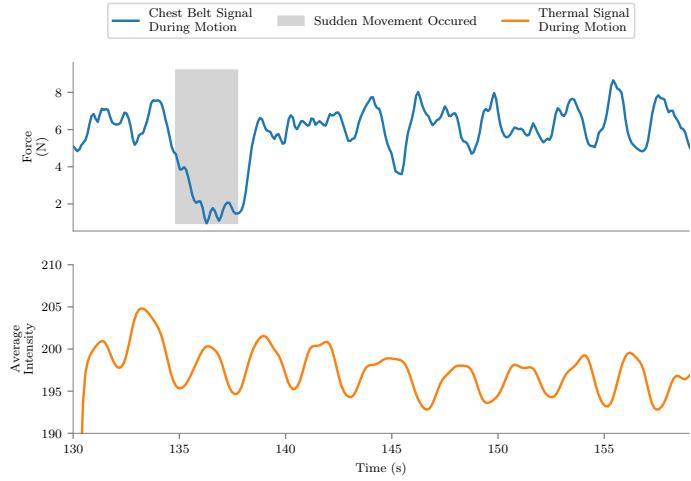


Fig. 3. In situations of vigorous motion, the chest belt signal demonstrates notably heightened noise levels compared to the respiration signal derived from thermal data. While the chest belt software registers a respiration rate of 5.4 breaths per minute, the simultaneous thermal data records a respiration rate of 22.2 breaths per minute—closely aligning with the manually determined ground truth of 20.0 breaths per minute. The chest belt’s subpar respiration rate performance can be attributed to the highlighted gray segment, where noise amplifies significantly in the chest belt signal, while the thermal signal remains stable. Remarkably, this degradation aligns with the participant’s body readjustment on the cycling seat. Further insights into thermal signal respiration extraction can be found in Section 4.5

3.1.2 Respiration Sensing Using RGB camera: In the past, RGB cameras have been used to sense respiration rate [13, 63, 73, 75]. However, such methods were not evaluated for capturing respiration during vigorous motion activities like cycling or running. Another fundamental principle behind extracting the respiration rate from RGB videos lies in sensing subtle changes in skin color. One such technique is Eulerian Video Magnification (VM) [77]. VM can amplify the color intensity change in the skin induced by breathing. Our observations revealed that during treadmill running, the cyclical human motion frequency aligns with the respiration rate frequency. Prior work [31] reported a similar observation. During a person’s running motion on the treadmill, the external room light falls on various parts of their face as they cycle. Unfortunately, this presents a challenge when using the VM algorithm, as it inadvertently amplifies this external light due to its frequency similarity to the respiration frequency range. Consequently, the skin’s color changes caused by respiration are subdued, leading to a failure in accurately sensing respiration from RGB video during motion. In contrast, extracting respiration data from thermal cameras does not require complex modeling, as the thermal imaging reveals the respiration pattern in the nostrils due to radiative and convective heat transfer components [67] throughout the breathing cycle, independent of external lighting conditions.

3.1.3 Respiration Sensing Using Earables: The adoption of earables for monitoring respiration rate has gained considerable momentum. Kumar et al.[41] effectively utilized earables to detect respiration patterns post-motion rather than within motion. However, JoulesEye aims to estimate respiration rates during motion to accurately calculate energy expenditure (EE). A recent study[2] has explored a multimodal approach using earables, combining audio and motion data to alleviate motion artifacts in various daily scenarios such as reclining, listening to music, and sitting or standing. Notably, this investigation primarily focused on head movements typical of

routine activities. Conversely, during intense physical exercises, head movements exhibit distinctive patterns unlike those observed during typical daily routines. While prior research [64] suggests excluding noisy samples caused by head movements, this strategy might be impractical during high-intensity workouts. We contend that evaluating the effectiveness of multimodal techniques solely based on regular daily head motions might inadequately represent their performance during rigorous physical activities.

3.2 Energy Expenditure Estimation

Prior research extensively explored the energy expenditure capabilities of consumer-grade wearables or wearable activity trackers [69]. Existing approaches for estimating energy expenditure often involve detecting activity using accelerometers, followed by applying predictive equations [5]. However, these methods face errors at two levels: first in activity classification and second in the significant variability of predictive equations. Studies have demonstrated the substantial variability of predictive equations utilizing accelerometers as inputs [47]. Well-known consumer smart wearables like Apple Watch and Fitbit estimate energy expenditure output based on Photoplethysmography (PPG), but comparative studies [17] reveal notable accuracy differences across brands. Recently, researchers [54] achieved accurate energy expenditure estimation using both Inertial Measurement Unit (IMU) and electrocardiogram (ECG) data. However, the ECG sensor required multiple probes placed across various torso areas. Other studies [40, 70] have explored energy expenditure estimation methods using IMU sensors on different body parts, such as the thigh and shank. In contrast, Aoki et al. [7] employed lasers to measure chest contraction as a proxy for ventilation threshold during stationary biking. While related, their study did not delve into investigating the ventilation threshold-energy expenditure relationship, nor did it address how chest occlusion from loose clothing might influence outcomes. Subsequent research by the same authors replaced the laser system with a Microsoft Kinect sensor for mobility, but this approach proved susceptible to participant movement [8]. Thermal videos have historically been used to measure energy expenditure [26, 38], leveraging optical flow to track cycling, walking, and running motion (but not respiration sensing). However, this solution applies mainly to non-contact energy expenditure monitoring within confined spaces where a single participant walks or runs. Optical flow tracking encounters difficulties in scenarios with multiple participants, uncontrolled movements, and occlusion.

Previous research has predominantly focused on singular types of exercise, potentially introducing bias to algorithms designed for energy expenditure estimation. While thermal cameras have been explored for non-contact energy expenditure estimation, these efforts have been confined to tracking privacy-preserving cycling motion using optical flow. *JoulesEye*'s novel contribution lies in employing a thermal camera to detect respiration rates, enabling energy expenditure estimation for participants engaged in both running and cycling exercises. The respiration rate extraction algorithms we employ demonstrate robustness against motion induced by cycling or running and are comparably resilient to occlusion in contrast to prior work. Addressing another challenge in energy estimation research, our approach aligns with the recommendation of the Intelligent Health and Well-Being Network (INTERLIVE) [9], a collaborative European initiative that aims to establish best-practice recommendations for evaluating the validity of consumer wearables and smartphones in the realm of physical activity assessment.

4 APPROACH

Our goal is to determine how many calories a person has burned while exercising by measuring the respiration rate. The breathing or respiratory rate is detected as a result of the temperature fluctuations due to airflow in the nasal. The physical phenomenon is based on the radiative and convective heat transfer component during the breathing cycle, which results in a periodic increase and decrease of the temperature at the tissues around

the nasal cavity. These observable temperature fluctuations are quantifiable in a thermal video as pixel intensity variations of the nostrils ROI [67]. As was explained in Section 3, the respiration rate is correlated with Energy Expenditure (EE). In this section, we first describe the setup of our system and then the algorithm for obtaining the temperature and respiration rate.

4.1 Ground Truth Devices



Fig. 4. JoulesEye’s is composed of a thermal camera retrofitted in an iPhone as shown in a). JoulesEye can be used in a smartwatch as shown in b). The camera in b) is a low resolution (32x24) thermal camera. c) and e) show the ground truth data collection procedure with indirect calorimeter while running and biking. d) shows a screen grab from the indirect calorimeter recording the energy expenditure during an exercise session.

First, we discuss the components used for collected ground truth data followed by other components used in the design of JoulesEye.

4.1.1 Indirect Calorimeter. We used an indirect calorimeter to collect the ground truth of energy expenditure. The Fitmate [74] indirect calorimeter (Figure 4(a-e)) consists of the following main components.

- Oxygen Sensor: Measures the oxygen consumption and the carbon dioxide expulsion from the body. The concentration of oxygen and carbon dioxide is directly proportional to the energy expenditure.
- Flow Sensor: Measures the volume of air breathed in and out by the user.
- Microprocessor Unit: Analyzes the data from the sensors and calculates the energy expenditure based on proprietary algorithms.
- Display Screen: It is shown in Figure 4(d). It displays the results of the energy expenditure calculation, including the number of calories burned, in real-time.

- Mouthpiece or Mask: Attaches to the face and connects to the calorimeter, allowing the measurement of inhaled and exhaled air.

The Fitmate Pro [74] indirect calorimeter is used to collect VO_2 (volume of oxygen) data during sub-maximal as well as maximal exercise by measuring the volume of oxygen consumed and the volume of carbon dioxide produced. Submaximal exercise is performed at a level below the maximum capacity of an individual. During physical assessment, only sub-maximal exercises should be performed by participants in the absence of a clinical physician [23]. The volume of oxygen consumed during physical activity is proportional to the amount of energy being expended. The unit of VO_2 is expressed as ml/kg/min. We convert VO_2 from ml/kg/min to calories (kcal/min), by using a conversion factor of 5 kcal/L [37]. Therefore, kcal/min = $(\text{VO}_2 \text{ (ml/kg/min)} \times \text{body weight (kg)} \times 5 \text{ kcal/L}) / 1000$. Throughout this paper, we use kcal/min as the unit of energy expenditure since ml/kg/min gives us a measure of the rate at which oxygen is consumed by the body during physical activity and not EE. The sampling rate of energy expenditure data is 0.5 Hz.

4.1.2 Respiration Belt. The ground truth of respiration rate is available from the indirect calorimeter. We also collect the respiration rate, using the Vernier GoDirect [22] respiration belt (Figure 4(e)). We collected respiration rate from two sources because it is impossible to use the calorimeter and *JoulesEye* simultaneously. Thus, instead of comparing *JoulesEye* with the gold standard output of the calorimeter, we use the chest belt as the reference measurement of respiration. The belt consists of a flexible, stretchable material that is worn around the chest, and it contains a sensor that detects the pressure changes caused by breathing. It has a measurement range of 0-100 breaths per minute with an error of ± 1 breath per minute. It has a sampling rate of 0.1 Hz.

4.2 *JoulesEye* System

JoulesEye consists of a thermal camera to record the respiration rate of a person. The thermal camera (Figure 4(a, b)) is used to retrieve the estimated value of respiration rate, temperature and used to estimate the EE. We used a FLIR One Pro [24] smartphone attachment thermal imaging camera. To take thermal videos, the device needs to be attached to an iPhone and connected to the FLIR ONE mobile application. A user can select the video mode and start recording. The thermal video will be recorded in real-time, showing temperature differences and heat patterns in the scene. The camera has a sampling rate of 8.6 frames per second with a temperature range of -20°C to 120°C. The combined unit of the smartphone and the thermal camera was securely mounted on the handgrip of an ergometer or affixed near the display screen of a treadmill in order to capture thermal video data of the face. We also developed a wristband prototype *JoulesEye* as shown in Figure 4(b). The details of the wristband prototype are described later in Section 6.5

4.3 Data Collection

To evaluate and inform our approach, we created a dataset which consisted of the following data.

- Energy Expenditure (EE) data measured in ml/kg/min along with volume of exhaled air measured in l/min.
- Thermal video data collected at 8.6 frames per second using a thermal camera. Thermal data is used for estimating respiration rate and temperature of the face.
- Reference respiration rate data collected using a respiration belt and an indirect calorimeter.
- Heart rate data measured in beats per minute collected via an Apple Watch as well as using a chest belt.
- RGB video data collected using iPhone 11 in parallel with thermal video data.

Total participants (n)	54
Participants who performed cycling on ergometer	41
Participants who performed running on treadmill	13
Female (n, %)	24, (44.4%)
Age (in years) (mean, range)	28.4 (25-54)

Table 2. Demographic information for the participants

4.4 Study Design

All participants between 18 to 70 years of age without any prior heart ailment could become a participant in the study. In total, 54 volunteers participated in an approximately 45 minutes study session (Table 2). The entry survey consisted of a questionnaire where the participants self-declared their age, weight, sex, time of the last meal and recent illnesses. Our data collection method followed the best practice validation protocol mandated by the Network of Physical Activity Assessment (INTERLIVE) [9]. The participants were shown how to wear the indirect calorimeter mask. All participants participated in two back-to-back data collection sessions.

4.4.1 Effect of Environmental Temperature. The environmental temperature affects only the temperature data that we retrieved from the forehead (explained later in Section 4.6.1) and not the respiration. Before the start of every data collection, we ensured that the effect of environmental temperature was minimised. We collected data in an air-conditioned room which had a temperature of 76 degrees Fahrenheit. We ensure that the participant has acclimatized to the room temperature before the data collection begins. The acclimatization can be confirmed by observing the nose tip of the person in the thermal video. The nose tip is one of the most sensitive region to external temperature [34]. During cold climate (data collection was held during winter months of December-January), the blood vessel stretches due to decreased blood flow as a result of homeostatic mechanisms resulting in decreased nose tip temperature compared to other regions of the face. Once the participant is indoors, the nose temperature and the nearby skin temperature gradually attains uniformity and that is when we consider the participant to have been acclimatized to the room temperature. Participant acclimatization is critical if we plan to use the temperature data (beside respiration rate) as a predictor of energy expenditure.

The only thing required to construct a respiration signal is the change in temperature in the nasal cavity. Irrespective of the environmental temperature, the temperature at the tissue around the nasal cavity will be different due to homeostatic mechanisms of the human body [34]. This implies that under normal physiological and environmental conditions, the change in temperature in the nasal cavity will always be visible in the thermal video.

The first session did not involve the indirect calorimeter because the calorimeter's mask occludes the nostril of the user, making it impossible for the thermal camera to extract respiration rate. Therefore, in the first session, we only collected thermal data without the indirect calorimeter to extract the respiration rate from the nostrils. The reference respiration signal of the first session was from the respiration belt alone. In the second session, we collected ground truth energy expenditure (EE) data from the indirect calorimeter and respiration data from the belt as well as the indirect calorimeter.

4.4.2 Session 1: Data Collection using JoulesEye. Participants cycled on a stationary bike or ran on a treadmill for both the sessions. In the first session, the participant ran for three minutes at a high intensity (4-5 miles/h running and 2.5-3 miles/h cycling). We limited the high intensity session to three minutes keeping the participant's comfort in mind. Figure 5(a) shows a frame of the face during this session. The following data are collected during this session.

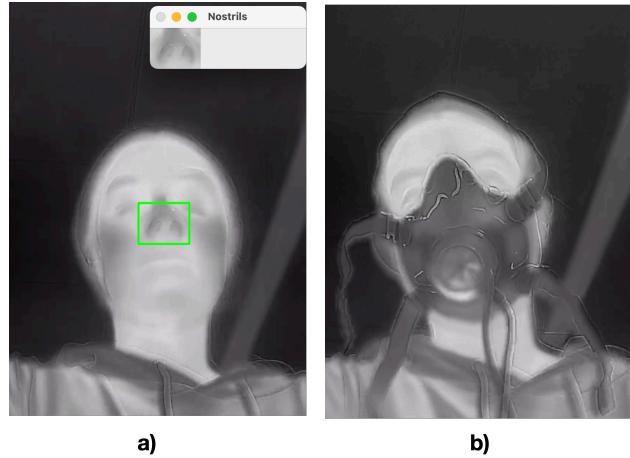


Fig. 5. In a) the participant has not donned the indirect calorimeter mask and hence the region tracking algorithm is able to keep track of the nostrils (nostril also shown in inset image). In b) the nostrils are covered by the mask making respiration detection impossible. We call a) as the *JoulesEye* data collection. During a) we could not collect the indirect calorimeter data in parallel as otherwise the nostrils would be occluded. Here, we use the respiration data from chest belt as the reference values. Thus, we could quantitatively evaluate the performance of *JoulesEye*'s respiration rate pipeline with ground truth respiration data from the belt. We later used the respiration rate from *JoulesEye* data to estimate energy expenditure

- Thermal video data of the upper body with the frame covering the face. This data is later processed to extract respiration rate.
- Respiration rate from the chest belt.

4.4.3 Session 2: Data Collection using Indirect Calorimeter. In this session, the participant donned the indirect calorimeter mask along with chest belt and performed cycling or running for 15 minutes comprising of High Intensity Interval Training (HIIT). We explain the details of HIIT in the next section (Section 4.4.4). The thermal camera recorded the face of the person during this session as well. Figure 5(b) shows a thermal frame of this session where the participant has donned the mask. Note that the nostrils are not visible and this thermal data cannot be used to extract respiration rate. The following data are collected during this session.

- Thermal data with frame covering the upper body including the face. The nostrils are now occluded by the mask.
- Respiration rate from the chest belt.
- Energy Expenditure, volume of exhaled air and respiration rate from the indirect calorimeter.

The respiration rate from the chest belt is a common modality between both the first and second sessions of data collection. We first evaluated our approach to estimating energy expenditure from respiration rate data from the indirect calorimeter. We next evaluated how of estimation of energy expenditure changes when we use respiration belt data. This evaluation helped us quantify the error of estimating energy expenditure from the belt's data. From the data collected in the first session, we quantified the error of estimating respiration rate via the thermal data in comparison to the chest belt. The idea is that if the respiration rate between all three modalities (Indirect Calorimeter, Belt and Thermal Data are within an acceptable range, we can claim that *JoulesEye* would be able to successfully estimate energy expenditure data from the respiration data alone.

We demonstrated in Section 4.2, that the chest belt data is not always reliable, especially during vigorous human motion. In those cases, we rely on our manual visual inspection of the thermal data for the ground truth. Section 4.5 details how respiration rate is extracted from the thermal data.

4.4.4 High Intensity Interval Training. In session two, the first 29 participants performed cycling at their own pace. The resistance of the ergometer was constant for these participants. During data collection we visually inspected the energy expenditure signal over time and realized that the energy expenditure gradually increases over time for all the participants. Our model might end up being biased to such a trend. To ensure diverse data in our dataset, we asked the next 25 participants to perform High Intensity Interval Training (HIIT) where we changed the resistance of the ergometer between highest and lowest level every 3 minutes. This lead to variation in the energy expenditure pattern and increases the diversity of EE data in our dataset. Participants, who participated in the treadmill running activity performed HIIT. They ran at 5-6 miles per hour during high intensity and 2.5-3 miles per hour during low intensity.

4.5 Extracting Respiration Rate During Motion From Thermal Video

The respiratory rate is identified by monitoring temperature changes in the airways caused by the flow of air. These noticeable temperature changes can be measured in a thermal video by observing variations in pixel intensity in the region of interest (ROI) encompassing the nostrils [67]. Given a correctly identified ROI, at time t , we computed the respiration signal, $I_{avg}(t)$ (respiration signal from infrared video) as

$$I_{avg}(t) = \frac{1}{WH} \sum_0^W \sum_0^H I_{ROI}(x, y, t)$$

where W and H are the width and height of the ROI, and $I_{ROI}(x, y, t)$ is the pixel intensity at pixel x, y at time t . (RR). We used 10 seconds of this signal ($I_{avg}(t)$) to compute a single sample of respiration rate (RR), details of which are in Section 5.

However, extracting physiological signals like respiration rate during physical activities, such as cycling and running is challenging due to significant body movement. We collected data in an uncontrolled environment where participants cycled and ran at a considerable pace (5-6 miles per hour), and their faces were recorded using a thermal camera. Uncontrolled environments bring in situations where participants would often rub off sweat using their hand (and occlude the camera) or momentarily move away from the frame of the camera. We used Channel and Spatial Reliability Tracker (CSRT) [46] algorithm to track the region of interest by accounting for such noises in the video. We found that the CSRT algorithm is better suited to track the region of interest (nostrils) in thermal video as compared to other state of the art landmark detection algorithms like MediaPipe [45] and OpenPose [14].

The CSRT algorithm uses both channel and spatial reliability information to track the target region. Channel reliability handles low contrast in thermal images, and spatial reliability handles variations in temperature and size of the target region. In terms of feature extraction, the algorithm extracts features from each color channel of the image separately. These features include a fusion of intensity and gradient information, such as Histogram of Oriented Gradients (HoG) and gray-scale intensity. To handle variations in the size and position of the target region, the algorithm also extracts spatial features like texture and shape from the image, which generate a spatial map that represents the likelihood of the target region being at each location in the image.

The CSRT is a region tracking algorithm which means that the region has to be identified manually before the algorithm starts tracking it. As discussed before, we are interested in tracking the nostril region and we use the RGB video data and classical landmark detection algorithm [45] to detect the nostrils in the initial frames spanning 10s of the video. Results demonstrate that the CSRT algorithm effectively tracks the nostril region throughout the exercise, even when some frames do not have the nostril visible due to occlusion. In Section 6.2.2 we discuss the effectiveness and limitations of the algorithm during occlusion. The algorithm maintains tracking of the region as long as the nostril returns in subsequent frames. Overall, the use of the CSRT algorithm allowed effective tracking of the nostril region during cycling and running activities, providing insights into physiological changes during exercise. An example video showing the nostril tracking can be seen in [this anonymized video](#).

4.6 Additional Data: Heart Rate and Temperature

We evaluated how temperature data and heart rate data can affect the energy expenditure estimation. We are interested in heart rate because (described in Section 3), heart rate is one of the most common proxies for energy expenditure and together with respiration rate, the estimations can improve. Given that that pixel intensity in the thermal data represents an approximate face temperature change during exercise, we can use this information as well to see the change in energy expenditure. The heart rate data collected via an Apple smart watch and the individual pixel intensity data collected from the thermal video was available for the second session. The following subsection details the approach on extracting temperature and heart rate.

4.6.1 Extracting Temperature: Our pilot experiments showed that temperature change occurs in the region of face with bony prominence like forehead, jawline and nose tip when a person is cycling an ergometer or running on the treadmill. It is known that physical activity increases the metabolic rate and generates heat in the body. This increased heat is transmitted through the blood vessels and nerves in the bony regions, leading to an increase in skin temperature in these regions [12]. We extracted temperature information from the forehead. We used the same region tracking algorithm with identical approach as described in Section 4.5, the only difference is that, this time, we track the forehead instead of the nostrils. We take the average intensity of the region of interest as a single sample or temperature value of a frame. Multiple frames across time give a 1D temperature vector.

4.6.2 Heart Rate: To make a fair comparison of the energy expenditure (EE) estimates produced by our approach, it was important to compare it with the currently accepted EE estimates produced by smart watches. The heart rate data from a Apple Watch was collected continuously during cycling and running, providing a continuous measurement of the individual's heart rate. This heart rate data was used as an additional optional input to our approach of estimate energy expenditure in combination with other physiological signals such as respiration rate and temperature. We also used the energy expenditure data from the Apple Watch as the reference for comparison with the energy expenditure estimates produced by our approach. By using the energy expenditure data from the Apple Watch as well as from our own model, it was possible to make a fair comparison of the accuracy of the energy expenditure estimates with respect to the ground truth.

4.7 Modeling

Energy Expenditure (EE) is represented in cal/min, deduced from VO_2 (Section 4.1.1). We aim to estimate Energy Expenditure (EE) from Respiration Rate (RR). We do this in two phases.

- We will first estimate the volume of exhaled air (v) from RR.
- Next, we will use the estimated volume information (v) to estimate the measures the oxygen concentration in a breath or VO_2 .

The inspiration of using this two-phased approach comes from the indirect calorimeter, which measures the oxygen concentration (O_2) in a breath. O_2 concentration in the inhaled air vary depending on factors like gas exchange efficiency and body composition. By trying to model the relationship between the amount of O_2 consumed and the volume of exhaled air (v), we can account for efficiency and body composition. But, v is not readily available without the indirect calorimeter. Therefore, our first objective is to estimate v from RR data. Using RR alone to estimate VO_2 can lead to inaccuracies because it does not take into account individual differences in lung capacities and breathing patterns. We expect our model to learn these factors to estimate v from RR alone. Our second model would then learn the transfer function and estimate unmeasured factors that would determine VO_2 from v .

4.7.1 Predicting volume from RR: Both our models are an adaptation of the Temporal Convolution Network with residuals (TCN) [10]. TCN leverages causal convolutions and dilation. Causal convolution enforces a unidirectional information flow, while dilation allows to capture long-range dependencies of the input. In our work, the model tries to learn a function f_1 that best predicts the volume v_t at time stamp t such that

$$v_t = f_1(v_{(t-k:t-1)}, RR_{(t-k:t)})$$

The model iterates over multiple samples of input and output to learn the function f_1 . During prediction, subsequent samples ($v_{(t+1)}, v_{(t+2)}\dots$), are predicted autoregressively i.e.

$$v_{(t+1)} = f_1(v_{(t-k+1:t)}, RR_{(t-k+1:t+1)})$$

where the predicted volume is used as an input to the next model which predicts VO_2 .

4.7.2 Predicting VO_2 from v : The approach to modeling VO_2 from volume is similar to the previous modeling approach where we use the TCN network, but this time we only use the volume information to predict VO_2 , i.e.

$$vo_{t+p} = f_2(v_t : v_{t+p-1})$$

where p is the number of samples of volume. Therefore, to predict the first sample of VO_2 , we need in total $k + p$ samples of respiration rate.

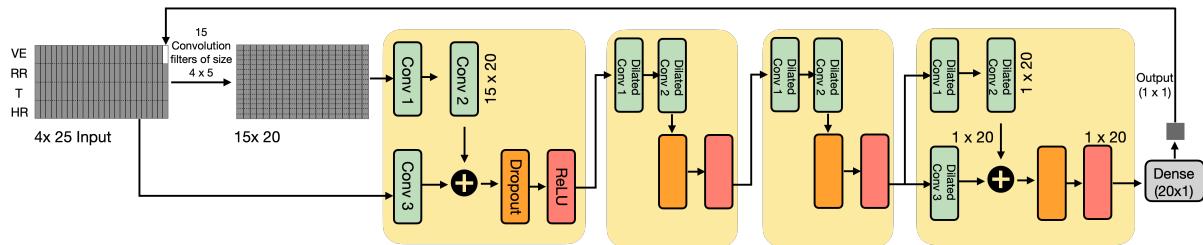


Fig. 6. We build a deep learning network similar to the Temporal Convolution Network (TCN) with residuals to estimate volume as a function of respiration rate and volume i.e. $v_t = f_1(v_{(t-k:t-1)}, RR_{(t-k:t)})$. Additionally, we also evaluated the performance of the model with additional covariates, namely heart rate (HR) and temperature (T) collected from the forehead. On using HR and T, the equation becomes, $v_t = f_1(v_{(t-k:t-1)}, HR_{(t-k:t)}, T_{(t-k:t)}, RR_{(t-k:t)})$. The residual blocks are composed of 1D dilated causal convolution (the first layer has no dilation), a ReLU activation and dropout [27]. A similar convolution is used to later predict VO_2 (calorie or energy expenditure) from Volume.

Using Heart Rate (HR) and Temperature Data (T): During data collection (Section 4.3), we retrieved heart rate data from both a chest belt and a smartwatch. Additionally, we obtained approximate temperature data from thermal readings. To enhance our analysis, we incorporated Heart Rate (HR) data from the smartwatch and forehead temperature (T) data as additional covariates. These supplementary variables enabled us to evaluate the performance of estimating v using different combinations of covariates, including HR alone, RR alone, a combination of RR and HR, and a combination of RR, HR, and T. For example, with an input of RR, HR and T, the equation to estimate v becomes

$$v_t = f_1(v_{(t-k:t-1)}, HR_{(t-k:t)}, T_{(t-k:t)}, RR_{(t-k:t)})$$

Figure 6 illustrates the corresponding TCN model for this combination of inputs. By modifying one (e.g., using only T and RR or T and HR) or two covariates (using only RR), we adjusted the input dimension of the model, necessitating corresponding adaptations to the kernel dimension while maintaining the dimensions of the tensors within the residual network unchanged.

5 EVALUATION

5.1 Experimental Setting For Data Collection

In this section we describe the internal settings of the three systems involved in collecting data (thermal camera, respiration belt and indirect calorimeter) and their synchronization. During data collection from the thermal camera, the device did not share raw temperature values and we had to extract temperature from pixel brightness from exported video. Thus, to counter the camera’s automatic gain control, we fixed its temperature range (via the FLIR mobile application) to 20–40°C. Fixing the temperature range simply ensures that the mapping between pixel intensity and temperature remains consistent for all the participants, *i.e.*, 20°C corresponds to 0 pixel intensity and 40°C corresponds to 255 pixel intensity for a grayscale video frame. This temperature range was found out experimentally as it is unlikely that a healthy human body would exceed 104°F (or 40°C) [28]. The respiration belt transmitted data via Bluetooth to a proprietary laptop application. The belt is also equipped with an on-device LED to indicate proper fit. The participant demographic information (age, weight, sex, and height) was inputted into the indirect calorimeter. Before each experiment, the indirect calorimeter underwent an internal calibration to assess the functionality of its flowmeter and the galvanic sensors.

We synchronized the data from the indirect calorimeter and the respiration belt. Both these machines were connected to a laptop during data collection which makes it possible to synchronize their timers. Similarly, the thermal camera attached to an iPhone reports a starting timestamp. The indirect calorimeter is enabled after the start of the thermal camera recording.

5.2 Experimental Setting For Extracting Respiration Rate and Temperature

We configured the belt to collect 10 samples of force data every second. In Section 4.5, we explained how CSRT algorithm tracks the nostrils to extract respiration rate. The region of interest is first identified using 10s of initial RGB video collected in parallel with thermal video. We used a classic landmark detection tool [45], to identify the nostrils as the region of interest. Thereafter, no RGB data is used and the CSRT tracker successfully keeps a track of the nostrils. The alignment between the RGB and the thermal data was achieved manually by adjusting the coordinate of the region of interest. Although the FLIR thermal camera has an inbuilt RGB camera, but it is redundant and we could not use this RGB data as raw data from the thermal camera is not readily available. Hence, we relied on the phone’s onboard camera for the RGB video.

The average pixel intensity of the region of interest over time gives us the respiration signal. The respiration signal has a frame rate of 8.6 samples per second (same as thermal video's fps). We extract one sample per frame. During post processing of the respiration signal, we do a linear interpolation post Gaussian smoothing such that we have 10 samples per second which matches with the sampling interval of the belt. Gaussian smoothing is better than a simple moving average since it does a better job of cutting off the higher frequencies [20]. To match the experimental settings of the respiration belt, we report the respiration rate every 3s from thermal data as well. Finally, we have the respiration from all three systems at a 3s interval. Similar to generating the 1D signal of respiration rate, we generate the 1D signal of the temperature. The only difference is that the region of interest is the forehead instead of the nostrils.

5.3 Experimental Setting For Modeling

We used leave-one-participant-out cross-validation (LOOCV), where the data from all but the test participant is used for training the model and evaluated on the test participant. LOOCV is preferred because it has less bias compared to the validation set approach where we randomly split the dataset into train/test/validation or use K-Fold cross validation [36]. We performed hyperparameter tuning on a random 40% of data from the training set for each of the folds. During preprocessing, we concatenate all participants' volume and RR data (additionally HR and T) to form a single 1D signal for both volume and RR, respectively. The 1D signal from all the participants is split into input and output chunks, where the input chunk represents data from the past, and the output chunk is what we want to predict. During training, the model takes as input the past data for volume and RR, and the current estimated RR to learn the current volume (See Figure 6). We ensured that all but the test participant forms the data of these 'chunks'. We performed hyperparameter tuning to find the optimal number of layers based on 'input chunk' size (l), kernel size (k), and dilation base (b) of the dilated convolution layer. These values are 20, 5 and 2, respectively. In TCN the number of layers (n) can be deduced using,

$$n = \lceil \log_b \left(\frac{(l-1)(b-1)}{(k-1)} \right) \rceil$$

which results into 3 layers [48]. The number of kernel filters was experimentally determined to be 15. We used a dropout of 0.03 with a learning rate of 0.005. We used early stopping to avoid overfitting and reduce training time. The early stopping ensures that the training stops if the validation mean absolute percentage error does not decrease by at least 0.05 for 5 consecutive epochs. The 'input chunk' size of 20 means that 20 values of respiration rate is required to estimate the first value of volume ($k = 20$ in Section 4.7). After which, every respiration rate sample returns a new sample of volume. 20 samples of respiration rate implies a time length of 60s. Therefore, 60s of respiration data is required for the first estimated sample of volume.

For the second TCN model that estimates VO_2 (and consequently the Energy Expenditure or EE) from volume of inhaled air data, the 'input chunk' of volume is 10 samples long ($p = 10$ in Section 4.7). This implies that to predict the first sample of VO_2 , $k + p = 20 + 10$ samples are required in total which finally boils down to 90s of data. Thus, 90s of input data is required for the first value of energy expenditure. Subsequent value of energy expenditure follows every second. We discuss the impact of changing the size of input chunk in Section 6.6. Additionally, we tested our model using different inputs such as respiration rate, heart rate from a consumer watch, a combination of respiration rate and temperature, a combination of respiration rate and heart rate, and a combination of respiration rate, heart rate, and temperature. We also tested our model on ground truth respiration rate and heart rate. Before feeding the data into the neural network, we changed the VO_2 data to cal/min data using the method mentioned in Section 4.1.1.

6 RESULTS AND DISCUSSION

In this section, we first discuss the performance of respiration detection from thermal video when compared to ground truth from a respiration belt. Next, we discuss the performance of energy expenditure estimation from respiration, heart rate, and temperature data.

6.1 Result on Estimating Respiration Rate

In this section, we will first discuss the result of our respiration rate estimation algorithm followed by the result on estimated energy expenditure (EE). With the data from the first session we observed that the error between respiration rate detection from thermal data when compared to respiration belt is 2.1% (Figure 7 (A)). Furthermore, from the data collected from the second session we quantified that the error between respiration rate detection from indirect calorimeter and respiration belt is 1.68%. In Section 3, we demonstrated that the belt tends to overestimate respiration rate in cases when it is not snugly fit or in situations when the participant performs sudden movement to readjust themselves especially when seated in the cycle. We meticulously annotated both the chest belt and thermal data. The visibility of respiration in the thermal video provides us with a means to rectify the chest belt data. Both these numbers (2.1% and 1.68%) are better compared to previous work [1] which uses Electrocardiogram and Photoplethysmogram to calculate respiration rate. Since both, respiration rate and energy expenditure are on different scales, using MAPE gives us a good idea of how changing one modality impacts the other.

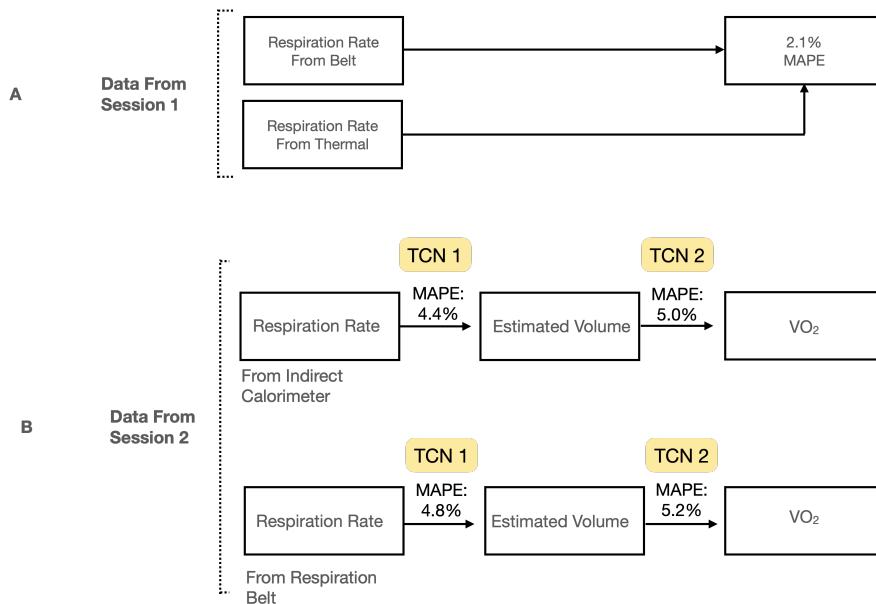


Fig. 7. The data obtained during the first session (**A**) serves the purpose of quantifying the discrepancy between the respiration signal extracted from the thermal video and the signal obtained from the belt. This quantification holds significance for subsequent insights, as depicted in Figure 8. The data collected during the second session (**B**) showcases the error in estimating VO_2 or EE when employing the respiration signal from an indirect calorimeter or a chest belt. Notably, it is important to recall that using an indirect calorimeter obstructs the view of the nostril from the thermal camera.

6.2 Result on Estimating Energy Expenditure

6.2.1 Using True and Reference Respiration Rate: Figure 7(B) shows the pipeline of estimating energy expenditure or VO_2 from ground truth respiration rate from the calorimeter and the reference respiration rate from the chest belt. The first TCN model is used to estimate volume of exhaled air from respiration rate data. The *estimated* volume of exhaled air is then used as an input to the second TCN model which estimates energy expenditure or VO_2 . Figure 9 shows that the best result of 5% Mean Absolute Percentage Error (MAPE) was obtained when ground truth respiration rate was input into the model. Using the belt's respiration rate as an input gives us an MAPE of 5.2%. To put these numbers into context, we compared the performance of using respiration rate as a predictor versus heart rate and temperature. We also compared the result obtained from Apple Smart Watch. In Figure 9, the following inputs are shown,

- True HR: This is the heart rate obtained from the indirect calorimeter chest band for heart rate.
- Estimated HR: This is the heart rate obtained using Apple Smartwatch.
- True RR: This is the respiration rate obtained from indirect calorimeter.
- Estimated RR: This is the RR generated by adding noise to the RR from the respiration belt. Recall that in Session 2, we collected respiration belt data. We add noise to this data so that we can simulate belt's data as the respiration data coming from thermal video (Explained in subsequent paragraph).
- Estimated RR and HR: This means the estimated RR data and Apple Watch HR data.
- Estimated RR, HR and T: This means the estimated RR data, Apple Watch HR data and the temperature data collected during session 2.

Figure 9 shows that the energy expenditure reported by Apple Watch has a MAPE of 37.5% across all participants when compared to the energy estimate from indirect calorimeter. But, using the heart rate from the Apple watch as an input to our TCN model gives an energy expenditure estimate with a MAPE of 12%. We believe this is because of two reasons. First, the algorithm that Apple Watch uses to estimate energy expenditure might be more generalised across exercises which are both anaerobic and aerobic and second, there might be some kind of exercise for which the energy expenditure algorithm of the watch expects hand movement to occur. For example, during cycling, there is no hand movement as participants usually rest their hands on the handrail.

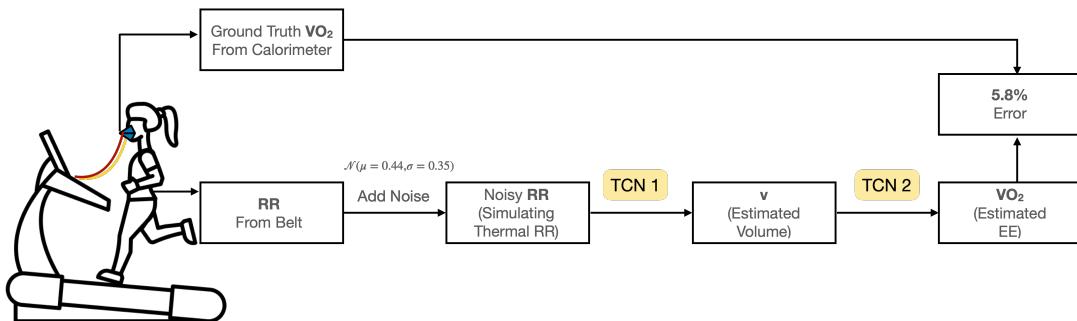


Fig. 8. JoulesEye EE estimation pipeline: The calorimeter's mask obstructs direct thermal-based respiration retrieval by blocking the camera's view of the nostrils. To replicate thermal-based respiration, we added noise to the belt-derived respiration signal, introducing a 2.1% error to simulate the difference between the reference respiration from the belt and the thermal video. The resulting noisy respiration rate (RR) signal was then input into the first TCN model for volume estimation. The estimated volume was subsequently passed to the second TCN model to predict VO_2 or energy expenditure.

6.2.2 Using Proxy Respiration Rate: We described in Section 4.3, why respiration rate obtained from thermal data is not available with ground truth of energy expenditure. But, from comparison of Session 1 data (Figure 7), we know that the error between respiration rate obtained from thermal data and belt is 2.1%. Thus, we can use the respiration data of belt from Session 2 (Figure 7) and add noise to it so that we generate new respiration data which has an error of 2.1%. We use this respiration data to predict energy expenditure as shown in Figure 8. Mathematically,

$$\text{NoisyRR}_i = \text{BeltRR}_i + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(\mu = 0.44, \sigma = 0.35)$$

where NoisyRR is the noisy respiration data and BeltRR is the respiration rate from the belt. We refer to this noisy respiration data as a proxy to the estimated respiration data. The choice of mean and standard deviation was such that so that the error between the noisy respiration rate and the ground truth respiration rate is 2.1%. A quantile-quantile probability plot confirmed that the respiration data from the belt follows a normal distribution and hence we choose the generate the noise from a normal distribution.

We compare the estimates from Apple Watch heart rate data and the estimates from estimated respiration rate in Figure 10. For cycling activity, the energy expenditure estimated from the heart rate data are relatively inaccurate as apparent from the noisy data in the lower portion of Figure 10(a). The same trend is not observed in Figure 10(b) when respiration rate is used as a predictor. Figure 9 also shows that adding Apple Heart Rate data to estimated respiration rate data can improve the performance from 5.8% to 5.5%. Adding temperature information improves the model further. But none of these improvements are as good as using the gold standard respiration rate from indirect calorimeter which has a MAPE of 5%. It is important to note that, for each participant, we changed the demographic information before data collection in the Apple Health app.

According to existing literature [53, 62], consumer grade health trackers should not exceed a 10.79% (MAPE) from the gold standard in order to be considered accurate. A previous evaluation of five commercially available wrist-worn devices tested with regard to their validity of EE compared with indirect calorimetry showed an MAPE > 10% [62]. Some other studies [68] put the MAPE at >20%. Our analysis has shown that heart rate from wrist worn wearable will not accurately predict EE (12%) but combining heart rate with respiration and temperature data will improve the MAPE to 5.2%.

6.3 Effect of occlusion

In Section 4.5, we asserted that the CSRT algorithm is capable of tolerating occlusion. To validate this claim, we conducted an evaluation to quantify the impact of occlusion on the accuracy of respiration rate estimation. Figure 11 (a) demonstrates that when three or more frames are consecutively occluded, the respiration rate estimation exhibits a high Mean Absolute Error of 20.1, while one or two frame occlusions result in a significantly lower error rate. The occurrence of prolonged occlusion, lasting three or more frames, leads to the loss of nostril tracking by the Region of Interest (ROI) tracker, causing alterations in the mean intensity signal, as illustrated in Figure 11 (b). As a consequence, this deviation in the intensity signal adversely affects the accuracy of respiration estimation. However, in such instances, we activate the RGB camera to re-establish the tracking of nostrils through landmark detection. By continuously tracking the nostril with the RGB camera, we can successfully retrieve the correct mean intensity signal, as shown in Figure 11 (c).

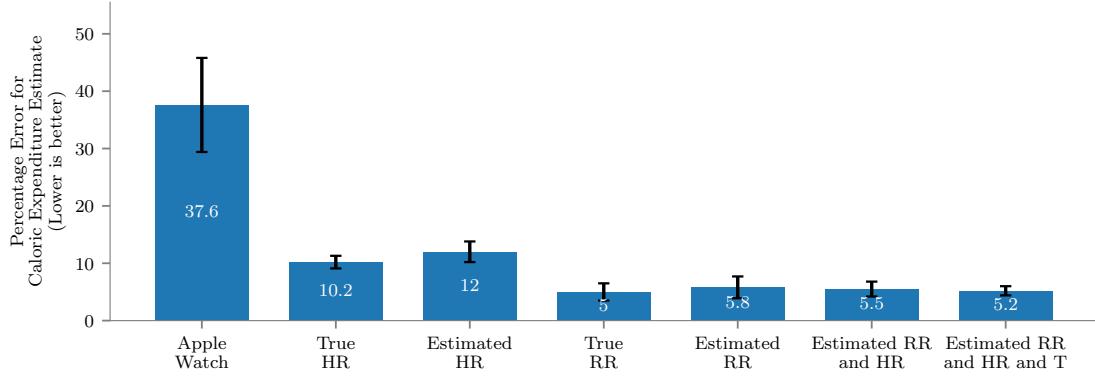


Fig. 9. **True HR** is the heart rate obtained from the indirect calorimeter chest band for heart rate. **Estimated HR** is the heart rate obtained using Apple Smartwatch. **True RR** is the respiration rate obtained from indirect calorimeter. **Estimated RR** is the JoulesEye's respiration rate. **Estimated RR and HR** is the estimated RR data and Apple Watch HR data and lastly the **Estimated RR, HR and T** is the estimated RR data, Apple Watch HR data and the temperature data collected during session 2 of data collection. Using respiration rate and temperature from JoulesEye and heart rate from smartwatch gives us MAPE of 5.2% which is much better compared to estimates from heart rate data alone (MAPE 10.2%).

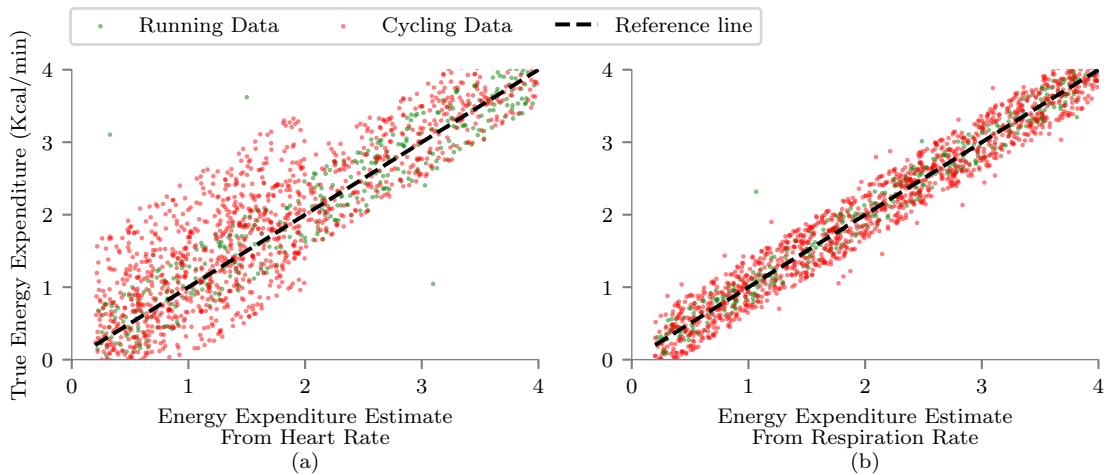


Fig. 10. In a), the error is higher for cycling activity, especially at lower energy expenditure rates. At low intensity cycling (2-2.5 Miles per hour), the Apple watch overestimates the heart rate. Although, the same is not true during running activity. In b), respiration rate does better in estimating both kind of exercise comprising high and low intensity exercise. Note that for visual clarity we have plotted a subset of samples from each participant (though the same trends exist on the whole dataset).

The data presented in Figure 11 (a) is the result of manual analysis conducted on a total of 2610 frames obtained from five participants. The occurrence of occlusion is predominantly observed when participants use their hands to rub their faces. As an example, Figure 11 (b) and (c) display a sequence of 526 frames from one participant. At the 250th frame (22nd second), the ROI tracker loses track of the nostril, prompting the manual activation

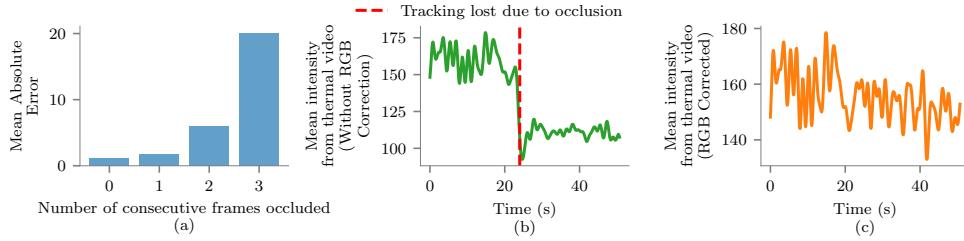


Fig. 11. The impact of occlusion: Analysis reveals that occlusion in three or more consecutive frames leads to a significant error in respiration estimation as shown in (a). Most of the occlusion occurs for 1 or 2 frame only, not leading to any significant error. **Nostril tracking Loss:** Prolonged occlusion disrupts the tracking of the nostril, resulting in the loss of the respiration signal as shown in (b). **Leveraging RGB camera for nostril reinstatement:** To address the issue of vigorous movement leading to nostril tracking loss, we introduce an RGB camera placed alongside the thermal camera. The RGB camera performs landmark detection on the nostril, reestablishing the nostril (ROI) tracker and preventing the loss of respiration signal as shown in (c).

of the RGB video at the 251st frame to detect the nostril. Once the nostril is successfully detected in the RGB footage, the ROI of the nostril is reinstated, ensuring the continuity of the tracking process. Further details on the matching process between RGB and thermal videos are explained in Section 4.5.

6.4 Discussion

Energy Expenditure (EE), or calories burned, is a complex process that varies among individuals due to a multitude of factors such as enzyme levels, gut bacteria, and body composition. Indirect calorimeters, considered the gold standard, measure heart rate and the concentration of inhaled oxygen and exhaled carbon dioxide to accurately estimate EE. On the other hand, consumer wearables worn on the wrist rely solely on heart rate for EE estimation. In our system, we enhance EE estimation by incorporating respiration rate and temperature alongside heart rate, resulting in improved performance, though not as precise as an indirect calorimeter. This advancement in EE estimation, particularly relevant in sports medicine, offers notable benefits. For instance, the International Society of Sports Nutrition [39] (ISSN) suggests that athletes engaging in intense training for 2–6 hours a day, 5–6 days a week, could burn over 1200 calories per hour during exercise. To maintain body weight, a corresponding calorie intake is necessary. An error exceeding 30% (as observed for smartwatches in Section 6.2) in EE estimation could lead to incorrect caloric consumption, resulting in suboptimal performance. As evident from Section 6.2, the inclusion of respiration rate in calorie estimation models substantially reduces error, facilitating more accurate post-exercise body weight management.

We will now provide an explanation of why respiration rate could be a better predictor for EE compared to HR with respect to our analysis and previous literature. According to literature [43, 44, 72], respiration rate information helps in explaining body composition or adiposity which is an important determiner for EE. Body composition plays a significant role in determining energy expenditure because each type of tissue in the body requires a different amount of energy to maintain. Muscle tissue is more metabolically active than fat tissue, meaning it requires more energy to sustain. This means that individuals with a higher proportion of muscle tissue will have a higher baseline energy expenditure compared to those with a higher proportion of fat tissue. People with a higher proportion of muscle tissue are generally able to perform physical activity more easily and for longer periods of time, leading to higher energy expenditure. On the other hand, individuals with a higher proportion of fat tissue may find physical activity more challenging and therefore burn fewer calories [76].

Adiposity, or the state of being overweight or obese, is closely related to body composition. Adipose tissue, or body fat, is a major component of body composition and contributes significantly to body weight. An increase in body fat is associated with an increase in respiration rate, as the body has to work harder to maintain its metabolic processes [44].

	All Participants	Participants With Normal BMI	Participants With Overweight BMI
Error (Apple Watch)	37.6%	29.7%	51.8%
Error (JoulesEye) with RR	5.8%	5.2%	6.9%

Table 3. We found that the Energy Expenditure estimates by Apple Watch is higher for people with high Body Mass Index (BMI), whereas it is relatively better for people with normal BMI.

To analyse if body composition affects the energy expenditure estimates, we split our data into people with normal and overweight Body Mass Index (BMI). Figure 3 shows the MAPE of energy expenditure from Apple Watch and the MAPE of energy expenditure estimates from respiration rate. It can be observed that the Apple Watch does a poorer job on estimating the correct energy expenditure for people who are overweight (MAPE: 51.8%) as compared to people who have normal BMI. Although, even respiration rate predicts energy expenditure better for people with normal BMI, but the MAPE is much lower, which might explain the theoretical understanding that respiration rate data helps in explaining the adiposity or the state of being overweight or obese.

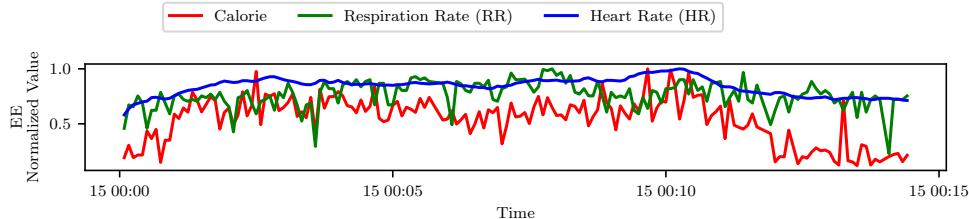


Fig. 12. Although, both heart rate and respiration rate are correlated with EE, using respiration rate helps explain the high frequency change in the EE signal.

Another reason why respiration rate explains the change in EE can be deduced from Figure 12 which shows that the heart rate, respiration rate and EE are well correlated, however the correlation between heart rate and energy expenditure is lower (Pearson Correlation = 0.78) compared to respiration rate and EE (0.93). Figure 12 suggests that high frequency information in the EE signal are captured by the respiration rate and not the heart rate. Heart rate signal is smoother resulting where no frequent changes are observed unlike in respiration rate and EE signal.

6.5 Result With Reduced Video Resolution From JoulesEye Smartwatch

As explained previously in Section 4.3, the FLIR Thermal camera needs to be retrofitted with an iPhone and its video recordings are saved in 1440x1080 pixel resolution without access to any raw data. But, for JoulesEye to be practical, we envision the smartwatch might come with a low resolution thermal camera. The primary advantage

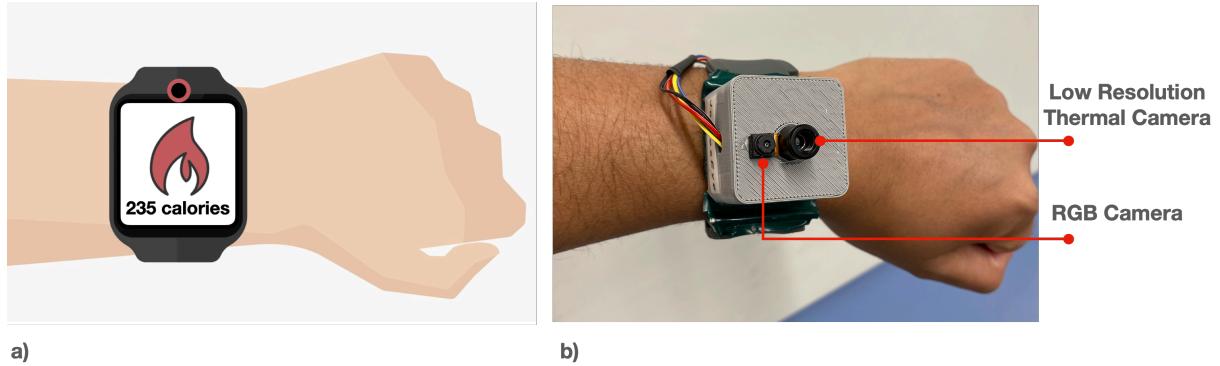


Fig. 13. We envision in the future, smartwatches can come installed with thermal cameras to accurately estimate EE. In a) we show what our future smartwatches can look like. In (b) we show our first prototype wristband thermal camera which is composed of a low resolution thermal camera and a RGB camera.

of using a low resolution thermal camera is reduced power and privacy concerns. As shown in Figure 13-b), we designed a 32x24 pixel resolution MLX90640 based thermal imaging system. It also has a RGB camera beside it. The RGB camera helps initially locate the nostrils and thereafter, the CSRT algorithm keeps track of the nostril as previously shown in Figure 2 in Section 1. We evaluated our low-resolution thermal system for respiration rate detection on 5 participants. These participants were asked to run on a treadmill at 4 miles per hour for a minute with the constraint that they look into the JoulesEye smartwatch thermal camera by extending their hand, akin to looking into a smartwatch. When compared to ground truth respiration rate data collected via the belt, we observed that the MAPE of estimating respiration rate is 8.1%. This high error arose because we were not able to achieve a high frame rate for the thermal camera. The current frame rate is 3 frames per second which is fine for slow or no movements but it causes a dithering effect when there is too much movement from the participant. We repeated the procedure as described in Section 6.2 of adding noise to respiration belt data so that the new data has an error of 8.1%. Using this data we got an energy expenditure estimate of 15.4%. While 15.4% is higher compared to the estimates from the watch's heart rate data alone (using our algorithms and not Apple Watch) which is 12% (Figure 9), combining this respiration rate data with heart rate data reduces the error to 10.1%. This shows that even though the frame rate of the wristband prototype is low, leveraging thermal data and heart rate data from smartwatch can estimate energy expenditure accurately when compared to heart rate data alone. The results are summarised in Figure 14.

We believe that a participant performing running or cycling need not always look into the smartwatch camera, but can periodically look into it so that instantaneous energy expenditure value can be calculated which in combination with heart rate data can improve the accuracy of energy expenditure as compared to heart rate alone. However, the minimum time required (which is 90s as described in Section 5.3) for estimating the initial value of energy expenditure poses a challenge. The next section, provides more insights into the time resolution of estimating energy expenditure.

6.6 Impact of Changing Time Resolution

Our result of JoulesEye shown in Figure 9 is based on input data sampling interval of 90s, where 60s is required to estimate the first sample of volume and further 30s more data is required to estimate the first sample of VO₂. Figure 15 shows how the percentage error changes when we gradually decrease the input chunk length

Resolution	Error on estimated RR
1080p thermal camera	2.1%
24p thermal camera	8.1%

Table 4: The error (MAPE) in RR estimation varies with changes in thermal video resolution.

Fig. 14. Estimation of EE using a low-resolution thermal camera in combination with heart rate data yields an error of 10.1%, showcasing its superiority over using heart rate data alone. These results demonstrate that even with a very low-resolution thermal camera, EE estimation can be enhanced. Addressing current challenges, such as dithering and time resolution (outlined in Section 6.6), holds the potential to further improve the accuracy of the estimation.

Input Data	Error on estimated EE
RR from 1080p thermal	5.4%
RR from 24p thermal	15.4%
RR from 1080p thermal and HR	5.3%
RR from 24p thermal and HR	10.1%

Table 5: Reduced thermal video resolution leads to increased error (MAPE) in EE estimation. However, the incorporation of HR data significantly enhances the results, even for low-resolution thermal videos.

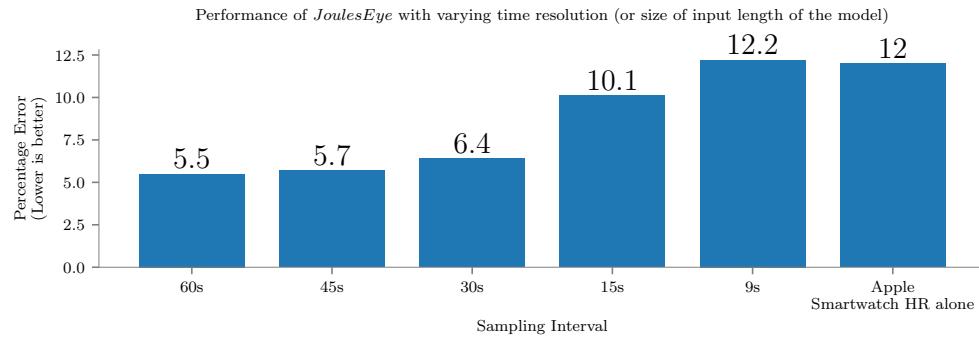


Fig. 15. Using 60s of respiration rate data gives us the best performance on estimation energy expenditure. 30s of respiration data is enough to predict energy expenditure with a better performance compared to heart rate alone.

of respiration rate estimation. We previously described about input chunk length in Section 5. We observe that using 15s of respiration data is enough to predict energy expenditure with a better performance as compared to heart rate alone. This implies that after exercising a user will have to look into the watch for 15s + 30s for her energy expenditure to be predicted by the model. We believe that work needs to be done in order to reduce this interval further towards making the system even more practical.

7 LIMITATIONS AND FUTURE WORK

We now discuss the limitations of our present work and plans for addressing them in the future.

- Smartphone/Smartwatch Integration: Our objective is to retrofit a smartphone/smartwatch with a low-resolution thermal camera [60] in line with our discussion in Section 6.5. Although we prototyped JoulesEye, engineering challenges to obtain higher frame rate remains an unsolved problem. Our initial result in Section 6.5 is promising, but our system is not yet real time, meaning the video processing and deep learning pipeline needs to be run after data recording.

- Usability of Smartwatch Prototype: Although we developed a prototype smartwatch for *JoulesEye*, we did not conduct any usability study with it. Currently, performing a usability study would not yield desired results, as each participant would need to continuously look into the watch for at least 45 seconds (as described in Section 6.6) to obtain any energy expenditure estimate. Such extended duration for glancing at the watch is impractical. Further research is required to significantly reduce this time interval, allowing a quick glance at the watch to provide accurate energy expenditure values.
- Uncertainty in Estimation: Our current methods for estimating energy expenditure give a point estimate. In the future, we plan to incorporate uncertainty in our estimation. Incorporating such uncertainty will be particularly important as various sensing modalities will be affected differently owing to differences in external conditions. As an example, the algorithms for heart rate estimation will likely not suffer even when the surroundings are dark, but the algorithms to estimate nostril position from RGB will suffer. Thus, in the future, we plan to implement a principled uncertainty based approach, where uncertainties in the different parts of the pipeline (estimating respiration rate, temperature; estimating energy expenditure using machine learning model) are considered while estimating energy expenditure.

8 CONCLUSION

Measuring energy expenditure during exercise is an important measure of overall health. In this work, we presented our system *JoulesEye* to accurately estimate energy expenditure during cycling and running. Our system measured the respiration rate using a thermal camera during motion and then estimated the energy expenditure. We found that our system is more accurate than smartwatch-based methods for estimating energy expenditure. We developed a low-resolution thermal camera prototype of our system that can be retrofitted in existing smartphones and watches and thus help scale measuring energy expenditure during exercise.

9 ACKNOWLEDGMENTS

The authors would like to express their gratitude to the reviewers for their valuable and actionable feedback. Rishiraj received support through the Fulbright-Nehru Doctoral Research Fellowship (FNDR), a joint award from the United States Government and the Government of India, during the course of this research. Rishiraj is also supported by the Prime Minister's Research Fellowship (PMRF) awarded by the Government of India. The authors also acknowledge the support received from the School of Computer Science (SCS) at Carnegie Mellon University (CMU), United States.

REFERENCES

- [1] Ali Adami, Reza Boostani, Faezeh Marzbanrad, and Peter H Charlton. 2021. A new framework to estimate breathing rate from electrocardiogram, photoplethysmogram, and blood pressure signals. *IEEE Access* 9 (2021), 45832–45844.
- [2] Tousif Ahmed, Md Mahbubur Rahman, Ebrahim Nemati, Mohsin Yusuf Ahmed, Jilong Kuang, and Alex Jun Gao. 2023. Remote breathing rate tracking in stationary position using the motion and acoustic sensors of earables. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [3] Mohamed Ali, Ali Elsayed, Arnaldo Mendez, Yvon Savaria, and Mohamad Sawan. 2021. Contact and remote breathing rate monitoring techniques: A review. *IEEE Sensors Journal* 21, 13 (2021), 14569–14586.
- [4] Abdulkadir H Alkali, Reza Saatchi, Heather Elphick, and Derek Burke. 2017. Thermal image processing for real-time non-contact respiration rate monitoring. *IET Circuits, Devices & Systems* 11, 2 (2017), 142–148.
- [5] Juan A Alvarez-Garcia, Božidar Cvetković, and Mitja Luštrek. 2020. A survey on energy expenditure estimation using wearable devices. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–35.
- [6] Heba Aly and Moustafa Youssef. 2016. Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.
- [7] Hirooki Aoki, Shiro Ichimura, Satoru Kiyooka, and Kohji Koshiji. 2007. Non-contact measurement method of respiratory movement under pedal stroke motion. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 374–377.

- [8] Hirooki Aoki and Hidetoshi Nakamura. 2018. Non-contact respiration measurement during exercise tolerance test by using kinect sensor. *Sports* 6, 1 (2018), 23.
- [9] Rob Argent, Megan Hetherington-Rauth, Julie Stang, Jakob Tarp, Francisco B Ortega, Pablo Molina-Garcia, Moritz Schumann, Wilhelm Bloch, Sulin Cheng, Anders Grøntved, et al. 2022. Recommendations for determining the validity of consumer wearables and smartphones for the estimation of energy expenditure: Expert statement and checklist of the INTERLIVE network. *Sports Medicine* 52, 8 (2022), 1817–1832.
- [10] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).
- [11] Brinnae Bent, Benjamin A Goldstein, Warren A Kibbe, and Jessilyn P Dunn. 2020. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ digital medicine* 3, 1 (2020), 18.
- [12] Mathias Bonmarin and Frédérique-Anne Le Gal. 2016. Thermal imaging in dermatology. *Imaging in dermatology* (2016), 437–454.
- [13] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. 2013. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control* 8, 6 (2013), 568–574.
- [14] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7291–7299.
- [15] Carl J Caspersen, Kenneth E Powell, and Gregory M Christenson. 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports* 100, 2 (1985), 126.
- [16] Ronan Chauvin, Mathieu Hamel, Simon Brière, François Ferland, François Grondin, Dominic Létourneau, Michel Tousignant, and François Michaud. 2014. Contact-free respiration rate monitoring using a pan–tilt thermal camera for stationary bike telerehabilitation sessions. *IEEE Systems Journal* 10, 3 (2014), 1046–1055.
- [17] Enhad A Chowdhury, Max J Western, Thomas E Nightingale, Oliver J Peacock, and Dylan Thompson. 2017. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *Plos one* 12, 2 (2017), e0171720.
- [18] Tarquin Collis, Richard B Devereux, Mary J Roman, Giovannii de Simone, Jeun-Liang Yeh, Barbara V Howard, Richard R Fabsitz, and Thomas K Welty. 2001. Relations of stroke volume and cardiac output to body composition: the strong heart study. *Circulation* 103, 6 (2001), 820–825.
- [19] Lilian De Greef, Mayank Goel, Min Joon Seo, Eric C Larson, James W Stout, James A Taylor, and Shwetak N Patel. 2014. Bilicam: using mobile phones to monitor newborn jaundice. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 331–342.
- [20] Allen Downey. 2016. *Think DSP: digital signal processing in Python*. " O'Reilly Media, Inc.".
- [21] Peter Dükking, Laura Giessing, Marie Otilie Frenkel, Karsten Koehler, Hans-Christer Holmberg, Billy Sperlich, et al. 2020. Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: validation study. *JMIR mHealth and uHealth* 8, 5 (2020), e16716.
- [22] Vernier Science Education. 2021. Go Direct Respiration Belt. <https://www.vernier.com/product/go-direct-respiration-belt/>. [Online; accessed 6-February-2023].
- [23] Gerald F Fletcher, Gary J Balady, Ezra A Amsterdam, Bernard Chaitman, Robert Eckel, Jerome Fleg, Victor F Froelicher, Arthur S Leon, Illeana L Piña, Roxanne Rodney, et al. 2001. Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association. *Circulation* 104, 14 (2001), 1694–1740.
- [24] Teledyne Flir. 2018. FLIR ONE Pro. <https://www.flir.com/products/flir-one-pro/?vertical=condition+monitoring&segment=solutions>. [Online; accessed 6-February-2023].
- [25] Daniel Fuller, Emily Colwell, Jonathan Low, Kassia Orychock, Melissa Ann Tobin, Bo Simango, Richard Buote, Desiree Van Heerden, Hui Luan, Kimberley Cullen, et al. 2020. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR mHealth and uHealth* 8, 9 (2020), e18694.
- [26] Rikke Gade, Ryan Godsk Larsen, and Thomas B Moeslund. 2017. Measuring energy expenditure in sports by thermal video analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 131–138.
- [27] Alberto Gasparin, Slobodan Lukovic, and Cesare Alippi. 2022. Deep learning for time series forecasting: The electric load case. *CAAI Transactions on Intelligence Technology* 7, 1 (2022), 1–25.
- [28] Ivayla I Geneva, Brian Cuzzo, Tasaduq Fazili, and Waleed Javaid. 2019. Normal body temperature: a systematic review. In *Open forum infectious diseases*, Vol. 6. Oxford University Press US, ofz032.
- [29] Riddhi Das Gupta, Roshna Ramachandran, Padmanaban Venkatesan, Shajith Anoop, Mini Joseph, and Nihal Thomas. 2017. Indirect calorimetry: from bench to bedside. *Indian journal of endocrinology and metabolism* 21, 4 (2017), 594.
- [30] Ehsanollah Habibi, Habibollah Dehghan, Mohammad Moghiseh, and Akbar Hasanzadeh. 2014. Study of the relationship between the aerobic capacity (VO_2 max) and the rating of perceived exertion based on the measurement of heart beat in the metal industries Esfahan. *Journal of education and health promotion* 3 (2014).
- [31] Nick Hayward, Mahdi Shaban, James Badger, Isobel Jones, Yang Wei, Daniel Spencer, Stefania Isichei, Martin Knight, James Otto, Gurinder Rayat, et al. 2022. A capacitive sensor provides continuous and accurate respiratory rate monitoring for patients at rest and during

- exercise. *Journal of Clinical Monitoring and Computing* 36, 5 (2022), 1535–1546.
- [32] Javier Hernandez, Yin Li, James M Rehg, and Rosalind W Picard. 2014. Bioglass: Physiological parameter estimation using a head-mounted wearable device. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. IEEE, 55–58.
- [33] Andrew P Hills, Najat Mokhtar, and Nuala M Byrne. 2014. Assessment of physical activity and energy expenditure: an overview of objective measures. *Frontiers in nutrition* 1 (2014), 5.
- [34] Stephanos Ioannou, Vittorio Gallesse, and Arcangelo Merla. 2014. Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology* 51, 10 (2014), 951–963.
- [35] Bashima Islam, Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Md Mehedi Hasan, Viswam Nathan, Korosh Vatanparvar, Ebrahim Nemati, Jilong Kuang, and Jun Alex Gao. 2021. BreathTrack: detecting regular breathing phases from unannotated acoustic data captured by a smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–22.
- [36] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An introduction to statistical learning*. Vol. 112. Springer.
- [37] Jeffrey M. Janot. 2005. Calculating Caloric Expenditure. IDEA Fitness Journal. https://www.ideafit.com/wp-content/uploads/files/_archive/062005_calculatin.pdf
- [38] Martin Møller Jensen, Mathias Krogh Poulsen, Thiemo Alldieck, Ryan Godsk Larsen, Rikke Gade, Thomas B Moeslund, and Jesper Franch. 2016. Estimation of energy expenditure during treadmill exercise via thermal imaging. *Medicine and science in sports and exercise* 48, 12 (2016), 2571–2579.
- [39] Chad M Kerksick, Colin D Wilborn, Michael D Roberts, Abbie Smith-Ryan, Susan M Kleiner, Ralf Jäger, Rick Collins, Mathew Cooke, Jaci N Davis, Elfego Galvan, et al. 2018. ISSN exercise & sports nutrition review update: research & recommendations. *Journal of the international society of sports nutrition* 15, 1 (2018), 38.
- [40] Salik Ram Khanal, Dennis Paulino, Jaime Sampaio, Joao Barroso, Arsénio Reis, and Vitor Filipe. 2022. A Review on Computer Vision Technology for Physical Exercise Monitoring. *Algorithms* 15, 12 (2022), 444.
- [41] Agni Kumar, Vikramjit Mitra, Carolyn Oliver, Adeeti Ullal, Matt Biddulph, and Irida Mance. 2021. Estimating respiratory rate from breath audio obtained through wearable microphones. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 7310–7315.
- [42] Eric C Larson, Mayank Goel, Gaetano Borriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N Patel. 2012. SpiroSmart: using a microphone to measure lung function on a mobile phone. In *Proceedings of the 2012 ACM Conference on ubiquitous computing*, 280–289.
- [43] Jing Li, Shiyue Li, Ritchie J Feuers, Cynthia K Buffington, and George SM Cowan Jr. 2001. Influence of body fat distribution on oxygen uptake and pulmonary performance in morbidly obese females during exercise. *Respirology* 6, 1 (2001), 9–13.
- [44] Stephen W Littleton. 2012. Impact of obesity on respiratory function. *Respirology* 17, 1 (2012), 43–49.
- [45] Camillo Lugaressi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).
- [46] Alan Lukezic, Tomas Vojir, Luka Čehovin Zajc, Jiri Matas, and Matej Kristan. 2017. Discriminative correlation filter with channel and spatial reliability. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6309–6318.
- [47] Kate Lyden, Sarah L Kozey, John W Staudenmeyer, and Patty S Freedson. 2011. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European journal of applied physiology* 111, 2 (2011), 187–201.
- [48] Francesco Lässig. 2022. Temporal Convolutional Networks and Forecasting. Unit8. <https://unit8.com/resources/temporal-convolutional-networks-and-forecasting/>
- [49] Alex Mariakakis, Megan A Banks, Lauren Phillipi, Lei Yu, James Taylor, and Shwetak N Patel. 2017. Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–26.
- [50] Alex Mariakakis, Jacob Baudin, Eric Whitmire, Vardhman Mehta, Megan A Banks, Anthony Law, Lynn McGrath, and Shwetak N Patel. 2017. PupilScreen: using smartphones to assess traumatic brain injury. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.
- [51] Alex Mariakakis, Edward Wang, Shwetak Patel, and Mayank Goel. 2019. Challenges in realizing smartphone-based health sensing. *IEEE Pervasive Computing* 18, 2 (2019), 76–84.
- [52] Didace Ndahimana and Eun-Kyung Kim. 2017. Measurement methods for physical activity and energy expenditure: a review. *Clinical nutrition research* 6, 2 (2017), 68–80.
- [53] M Benjamin Nelson, Leonard A Kaminsky, D Clark Dickin, and ALEXANDER H Montoye. 2016. Validity of consumer-based physical activity monitors for specific activity types. *Medicine and science in sports and exercise* 48, 8 (2016), 1619–1628.
- [54] Zhiqiang Ni, Tongde Wu, Tao Wang, Fangmin Sun, and Ye Li. 2022. Deep multi-branch two-stage regression network for accurate energy expenditure estimation with ECG and IMU data. *IEEE Transactions on Biomedical Engineering* 69, 10 (2022), 3224–3233.
- [55] Andrea Nicolò, Michele Girardi, Ilenia Bazzucchi, Francesco Felici, and Massimo Sacchetti. 2018. Respiratory frequency and tidal volume during exercise: differential control and unbalanced interdependence. *Physiological reports* 6, 21 (2018), e13908.

- [56] Andrea Nicolò, Samuele M Marcora, and Massimo Sacchetti. 2016. Respiratory frequency is strongly associated with perceived exertion during time trials of different duration. *Journal of sports sciences* 34, 13 (2016), 1199–1206.
- [57] Andrea Nicolò and Massimo Sacchetti. 2019. A new model of ventilatory control during exercise. *Experimental physiology* 104, 9 (2019), 1331–1332.
- [58] World Health Organisation. 2021. Obesity and overweight. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>. [Online; accessed 13-January-2023].
- [59] Ruairí O'Driscoll, Jake Turicchi, Kristine Beaulieu, Sarah Scott, Jamie Matu, Kevin Deighton, Graham Finlayson, and James Stubbs. 2020. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine* 54, 6 (2020), 332–340.
- [60] Panasonic. 2017. GridEye Infinite Array Sensor. Panasonic Corporation of North America. <https://na.industrial.panasonic.com/products/sensors/sensors-automotive-industrial-applications/lineup/grid-eye-infrared-array-sensor>
- [61] Jonghoon Park, Ishikawa-Takata Kazuko, Eunkyoung Kim, Jeonghyun Kim, and Jinsook Yoon. 2014. Estimating free-living human energy expenditure: practical aspects of the doubly labeled water method and its applications. *Nutrition research and practice* 8, 3 (2014), 241–248.
- [62] Stefanie Passler, Julian Bohrer, Lukas Blöchinger, and Veit Senner. 2019. Validity of wrist-worn activity trackers for estimating VO_{2max} and energy expenditure. *International journal of environmental research and public health* 16, 17 (2019), 3037.
- [63] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. 2010. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering* 58, 1 (2010), 7–11.
- [64] Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Ebrahim Nemati, Minh Dinh, Nathan Folkman, Md Mehedi Hasan, Jilong Kuang, and Jun Alex Gao. 2021. Towards motion-aware passive resting respiratory rate monitoring using earbuds. In *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–4.
- [65] Polar Research and Technology. 2018. Polar Smart Calories. <https://www.polar.com/sites/default/files/static/science/white-papers/polar-smart-calories-white-paper.pdf>. [Online; accessed 17-January-2023].
- [66] Tobias Röddiger, Daniel Wolffram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards respiration rate monitoring using an in-ear headphone inertial measurement unit. In *Proceedings of the 1st International Workshop on Earable Computing*. 48–53.
- [67] Gaetano Scecca, Giulia Da Poian, and Walter Karlen. 2020. Multispectral video fusion for non-contact monitoring of respiratory rate and apnea. *IEEE Transactions on Biomedical Engineering* 68, 1 (2020), 350–359.
- [68] Anna Shcherbina, C Mikael Mattsson, Daryl Waggett, Heidi Salisbury, Jeffrey W Christle, Trevor Hastie, Matthew T Wheeler, and Euan A Ashley. 2017. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of personalized medicine* 7, 2 (2017), 3.
- [69] Ren-Jay Shei, Ian G Holder, Alicia S Oumsang, Brittni A Paris, and Hunter L Paris. 2022. Wearable activity trackers—advanced technology or advanced marketing? *European Journal of Applied Physiology* 122, 9 (2022), 1975–1990.
- [70] Patrick Slade, Mykel J Kochenderfer, Scott L Delp, and Steven H Collins. 2021. Sensing leg movement enhances wearable monitoring of energy expenditure. *Nature Communications* 12, 1 (2021), 4312.
- [71] Xiao Sun, Li Qiu, Yibo Wu, Yeming Tang, and Guohong Cao. 2017. Sleepmonitor: Monitoring respiratory rate and body position during sleep using smartwatch. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–22.
- [72] Tim JT Sutherland, Christene R McLachlan, Malcolm R Sears, Richie Poultney, and Robert J Hancox. 2016. The relationship between body fat and respiratory function in young adults. *European Respiratory Journal* 48, 3 (2016), 734–747.
- [73] Lionel Tarassenko, Mauricio Villarroel, Alessandro Guazzi, Joao Jorge, DA Clifton, and Chris Pugh. 2014. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological measurement* 35, 5 (2014), 807.
- [74] COSMED the metabolic company. 2008. Fitmate PRO. <https://www.cosmed.com/en/products/cardio-pulmonary-exercise-test/fitmate-pro>. [Online; accessed 6-February-2023].
- [75] Mark Van Gastel, Sander Stuijk, and Gerard de Haan. 2016. Robust respiration detection from remote photoplethysmography. *Biomedical optics express* 7, 12 (2016), 4941–4957.
- [76] Thomas A Wadden, Renee A Vogt, Ross E Andersen, Susan J Bartlett, Gary D Foster, Robert H Kuehnel, Joshua Wilk, Ruth Weinstock, Philip Buckenmeyer, Robert I Berkowitz, et al. 1997. Exercise in the treatment of obesity: effects of four interventions on body composition, resting energy expenditure, appetite, and mood. *Journal of consulting and clinical psychology* 65, 2 (1997), 269.
- [77] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)* 31, 4 (2012), 1–8.
- [78] Yanni Yang, Jiannong Cao, and Xiulong Liu. 2019. ER-rhythm: Coupling exercise and respiration rhythm using lightweight COTS RFID. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–24.
- [79] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dima Katabi. 2018. Extracting multi-person respiration from entangled RF signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–22.