

Article

# Adaptive Thermal Imaging Signal Analysis for Real-Time Non-Invasive Respiratory Rate Monitoring

Riska Analia <sup>1,2</sup> , Anne Forster <sup>3</sup> , Sheng-Quan Xie <sup>1</sup>  and Zhiqiang Zhang <sup>1,\*</sup> 

<sup>1</sup> School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9LN, UK; elran@leeds.ac.uk (R.A.); s.q.xie@leeds.ac.uk (S.-Q.X.)

<sup>2</sup> Department of Electrical Engineering, Politeknik Negeri Batam, Batam 29461, Indonesia

<sup>3</sup> Academic Unit for Ageing and Stroke Research, Leeds Institute of Health Sciences, University of Leeds, Leeds LS2 9JT, UK; a.forster@leeds.ac.uk

\* Correspondence: z.zhang3@leeds.ac.uk

## Abstract

(1) Background: This study presents an adaptive, contactless, and privacy-preserving respiratory-rate monitoring system based on thermal imaging, designed for real-time operation on embedded edge hardware. The system continuously processes temperature data from a compact thermal camera without external computation, enabling practical deployment for home or clinical vital-sign monitoring. (2) Methods: Thermal frames are captured using a  $256 \times 192$  TOPDON TC001 camera and processed entirely on an NVIDIA Jetson Orin Nano. A YOLO-based detector localizes the nostril region in every even frame (stride = 2) to reduce the computation load, while a Kalman filter predicts the ROI position on skipped frames to maintain spatial continuity and suppress motion jitter. From the stabilized ROI, a temperature-based breathing signal is extracted and analyzed through an adaptive median–MAD hysteresis algorithm that dynamically adjusts to signal amplitude and noise variations for breathing phase detection. Respiratory rate (*RR*) is computed from inter-breath intervals (IBI) validated within physiological constraints. (3) Results: Ten healthy subjects participated in six experimental conditions including resting, paced breathing, speech, off-axis yaw, posture (supine), and distance variations up to 2.0 m. Across these conditions, the system attained a MAE of  $0.57 \pm 0.36$  BPM and an RMSE of  $0.64 \pm 0.42$  BPM, demonstrating stable accuracy under motion and thermal drift. Compared with peak-based and FFT spectral baselines, the proposed method reduced errors by a large margin across all conditions. (4) Conclusions: The findings confirm that accurate and robust respiratory-rate estimation can be achieved using a low-resolution thermal sensor running entirely on an embedded edge device. The combination of YOLO-based nostril detector, Kalman ROI prediction, and adaptive MAD–hysteresis phase that self-adjusts to signal variability provides a compact, efficient, and privacy-preserving solution for non-invasive vital-sign monitoring in real-world environments.



Academic Editor: Jianxiong Zhu

Received: 20 November 2025

Revised: 12 December 2025

Accepted: 25 December 2025

Published: 1 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

**Keywords:** adaptive; contactless; embedded edge hardware; inter-breath intervals; non-invasive sensing; respiratory rate monitoring; real-time; thermal imaging

## 1. Introduction

Continuous monitoring of respiratory rate (*RR*) has become one of the most important vital signs in both clinical and home care settings. *RR* plays a crucial role in assessing a physiological condition of the patient, especially during clinical deterioration [1–3]. Early detection through continuous *RR* tracking enables timely intervention, particularly for

high-risk populations [4–7]. In long-lie conditions following a fall, continuous respiratory monitoring offers valuable physiological information for early detection and timely assistance [8–10].

Several conventional approaches have been employed in respiratory monitoring studies. These typically involve physical contact with the patient, including chest bands, nasal cannulas, or spirometry devices. While these tools are clinically validated, they often cause discomfort and may interfere with natural breathing behavior, especially during sleep or prolonged observation periods [11–14]. A non-contact alternative method, on the other hand, has shown promising results in estimating  $RR$ , which could be one of the solutions to overcome these challenges. A recent study developed a non-contact system using radar-based techniques, acoustic sensors, and camera-based methods such as RGB or thermal imaging [15–18].

Among these non-contact modalities, thermal imaging presents distinct advantages for  $RR$  estimation. It enables the capture of temperature variations generated by inhaled and exhaled air without physical contact. These thermal fluctuations, observed around the nostril or mouth region, offer a natural and unobtrusive signal source for respiratory analysis [19,20]. This makes thermal-based systems particularly well-suited for continuous monitoring in privacy-sensitive environments such as bedrooms or elder care facilities.

Despite its promise, thermal-based respiratory monitoring still faces several technical challenges. Accurate detection and tracking of the nostril region is often hindered by the low spatial resolution of thermal cameras, which complicates region of interest (ROI) localization [21,22]. Furthermore, thermal signals are susceptible to noise introduced by subject movement, head rotations, and ambient temperature changes, all of which can degrade signal quality and affect  $RR$  estimation accuracy [22,23]. Moreover, many existing implementations rely on frequency-domain analysis or computationally expensive deep learning models, which limits their real-time feasibility on embedded platforms [19,24–27].

Recent biomedical monitoring systems have increasingly shifted toward embedded edge computing due to its advantages in latency, privacy, and deployment feasibility. Unlike cloud-based processing, which introduces transmission delays and raises concerns over sensitive health data exposure, edge computation allows all inference to occur locally on the device. This enables real-time responsiveness and preserves subject privacy, which are essential for continuous respiratory monitoring [28–32]. However, most existing thermal and camera-based respiratory monitoring studies still rely on offline processing pipelines due to their high computational requirements, limiting their applicability for real-time embedded deployment.

Therefore, current thermal-based  $RR$  approaches still leave three critical gaps unaddressed: (i) the lack of a robust nostril-specific localization strategy for low-resolution thermal imagery, resulting in unstable ROI tracking; (ii) the absence of computational optimization needed for real-time deployment on embedded edge devices; and (iii) limited robustness of time-domain phase detection, which remains sensitive to motion-induced disturbances, amplitude variability, and thermal drift. These unmet needs motivate the development of a thermal-specific, computation-efficient, and motion-resilient respiratory monitoring framework suitable for continuous operation in real-world environments.

To overcome these limitations, this study introduces a fully automated, privacy-preserving thermal-imaging system for real-time respiratory-rate monitoring on embedded edge hardware. The system begins with a thermal YOLO-based model to locate the nostril region as a small-object bounding box; this box defines the region of interest (ROI) from which an airflow-related temperature signal is extracted using the coldest pixel within the ROI, reflecting inhalation–exhalation temperature modulation. To reduce computational cost at the detector stage, an adaptive frame skipping (stride = 2) with Kalman prediction

is applied so that the YOLO detector runs at half the nominal frequency while a Kalman tracker tracks the nostril bounding-box (bbox) between detections, preserving continuity and suppressing motion artifacts. The YOLO-based nostril detector was implemented using the YOLOv8n model, which naturally supports small-object detection and operates efficiently on low-resolution thermal imagery.

Respiratory-rate estimation begins with breathing-phase detection on the stabilized ROI signal using an adaptive hysteresis state machine driven by velocity-based thresholds. These thresholds are derived from the median absolute deviation (MAD) and integrated with a flicker-suppression mechanism to maintain signal stability during head movement and other motion disturbances. The resulting stable breathing-phase sequence is then used to determine inter-breath intervals (IBI), from which the respiratory rate (RR) is calculated.

The main contributions of this work are: (i) a thermal-specific YOLO-based nostril detector designed for small-object detection in low-resolution  $256 \times 192$  thermal imagery, overcoming ROI instability common in prior thermal RR studies, (ii) a detector-centric frame-skipping mechanism (stride = 2) integrated with Kalman ROI prediction, reducing detection computation by 50% while maintaining spatial continuity and enabling real-time embedded operation, (iii) an adaptive time-domain respiratory-phase detection approach that combines median-MAD thresholds, hysteresis, and flicker suppression to achieve robust segmentation under motion, drift, and amplitude variability, without relying on frequency-domain analysis, (iv) a fully on-device respiratory-rate monitoring pipeline, running entirely on an NVIDIA Jetson Orin Nano without cloud processing, ensuring privacy preservation and practical deployment feasibility for long-term ambient monitoring, and (v) a comprehensive evaluation across six real-world conditions (resting, paced breathing, soft speech, off-axis yaw, distance variation up to 2.0 m, and supine posture), demonstrating clinically acceptable accuracy (overall MAE  $0.57 \pm 0.36$  BPM), outperforming previously reported thermal-based contactless RR systems. To guide the development of this work, the following research questions are formulated:

RQ1: Can a low-resolution thermal camera, combined with automated nostril tracking, provide accurate and reliable respiratory-rate estimation across diverse real-world conditions?

RQ2: How can adaptive signal-processing strategies, such as MAD-based breathing-phase detection and IBI validation, improve robustness against facial movement, off-axis orientation, and varying thermal contrast?

RQ3: Is the proposed approach computationally lightweight enough to operate in real time on an embedded edge device without compromising accuracy?

These questions motivate the system design and experimental evaluation presented in the remainder of this paper.

The remainder of this paper is structured as follows: related work is reviewed in Section 2; Section 3 describes the proposed methodology, including data acquisition and model architecture; Section 4 presents experimental results; Section 5 discusses system performance and deployment feasibility; and Section 6 concludes the paper with a summary and directions for future work.

## 2. Related Work

Among the human vital signs, respiratory rate (RR) is widely recognized as a critical indicator of physiological stability. In estimating the RR, thermal imaging has emerged as a promising non-contact and privacy-preserving modality. Unlike contact-based methods that require direct attachment to the body, thermal cameras detect temperature variations caused by airflow during inhalation and exhalation, typically around the nostrils or mouth. These thermal fluctuations form a natural and unobtrusive signal source for respiratory

analysis, particularly suitable for continuous monitoring in both clinical and home care settings [33–35].

A variety of methods have been proposed to extract respiratory signals from thermal video data. Earlier approaches relied on manual region-of-interest (ROI) selection and simple pixel averaging, which were limited robustness under motion or occlusion. More recent developments introduced computer vision and deep learning techniques for automated ROI localization, including three-dimensional convolutional neural networks, detection transformers, and single-shot detectors such as YOLO and SSD [26,27,36]. Facial landmark-based approaches have also been used to improve stability under head motion or partial occlusion [37,38]. Other studies align RGB landmarks with thermal images [22] or extract thermal-motion data to detect breathing regions even when facial features are obscured by masks or bedding [23,39,40]. However, small-object detection of the nostril region in low-resolution thermal frames remains challenging.

Once the ROI is detected respiratory signals are typically obtained by tracking temperature variations over time. To enhance signal quality, various filtering methods such as Butterworth, Hampel, and Savitzky–Golay have been employed, along with adaptive decomposition techniques like the Hilbert–Huang Transform [36,41,42]. Other preprocessing strategies, including histogram equalization, optimal quantitation, and super-resolution, have been applied to compensate for the low resolution of compact thermal cameras [43]. Moreover, some works directly apply deep models to thermal sequences, learning temporal breathing patterns without explicit ROI tracking [19,26]. Many studies estimate RR using dominant spectral components via Fourier, synchro-squeezed, or autocorrelation analysis [27,34,42], but frequency-domain approaches can be sensitive to noise, motion artifacts, and ambient temperature drift. Few works incorporate explicit flicker-suppression or median-absolute-deviation-based adaptive thresholds in the breathing phase logic to stabilize the signal during head movement.

Beyond algorithmic advances, several recent works emphasize that embedded edge computing has become a key requirement for biomedical sensing systems [28,29]. On-device processing reduces communication overhead, enhances data security, and enables deployment in resource-constrained settings where continuous cloud connectivity is impractical [30–32]. Despite the increasing adoption of edge-based architectures, existing thermal RR methods seldom address the computational constraints of embedded hardware, with many relying on high-resolution cameras or offline deep models. This gap further motivates the development of a lightweight and fully embedded respiratory-rate monitoring pipeline.

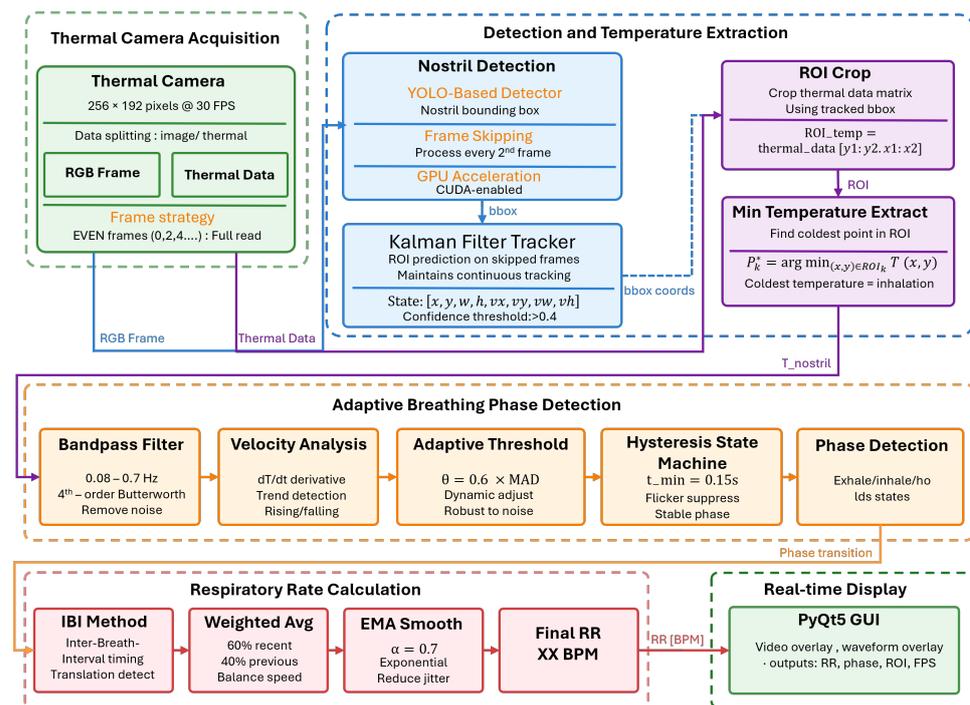
Based on the gaps outlined above, this paper proposes an adaptive, real-time thermal respiratory monitoring system for embedded deployment. The contributions comprise a thermal-specific YOLO-based detector for nostril localization, a detector-stage frame-skipping scheme with Kalman prediction to halve detection frequency while preserving ROI continuity, and an adaptive MAD-hysteresis phase detection framework with flicker suppression for motion-robust, physiologically consistent respiratory-rate estimation, all validated on-device under privacy-preserving constraints.

### 3. Materials and Methods

#### 3.1. System Overview

The proposed system converts raw thermal video frames into RR estimation through a sequence of tightly coupled modules, as illustrated in Figure 1. Thermal frames are first captured by a low-resolution thermal camera that provides two synchronous outputs: an thermal imagery frame and a thermal data. A YOLO-based detector localizes the nostril on the thermal imagery every second frame, and the resulting ROI is projected to the

thermal mapping image; between detections, a Kalman prediction updates the ROI directly in thermal coordinates. The thermal data are decoded into per-pixel temperatures to form a calibrated temperature map; the tracked bounding box crops this map, and the coldest pixel temperature per frame serves as a one-dimensional airflow-related signal. This signal is band-pass filtered in the 0.08 to 0.7 Hz range with a fourth-order zero-phase Butterworth filter and analyzed in the time domain using velocity estimates with median-absolute-deviation (MAD)-derived thresholds. An adaptive hysteresis state machine with a minimum dwell of 0.15 s produces inhale, exhale, and hold phases. Phase transitions yield inter-breath intervals (IBI) that are validated within a physiologic range and converted to breaths per minute (BPM), then stabilized by short weighted averaging and exponential moving averaging.



**Figure 1.** System overview of the proposed real-time thermal-based respiratory rate estimation. YOLO-based detector with Kalman tracking stabilizes the nostril region of interest (ROI), from which the coldest pixel temperature is extracted as the airflow-related signal. ROI min-temperature is band-pass filtered and analysed with MAD-based hysteresis and IBI validation to produce respiratory rate.

### 3.2. Thermal Camera Acquisition

The initial captured frame from a thermal camera consists of both image and thermal information; therefore, splitting image and thermal information is necessary as the first step in processing thermal camera acquisition data. The process of thermal acquisition is illustrated in Figure 2. The initial capture frame can be expressed as

$$I_{\text{raw}} \in \mathbb{Z}_{256}^{H \times 2W}, \quad (1)$$

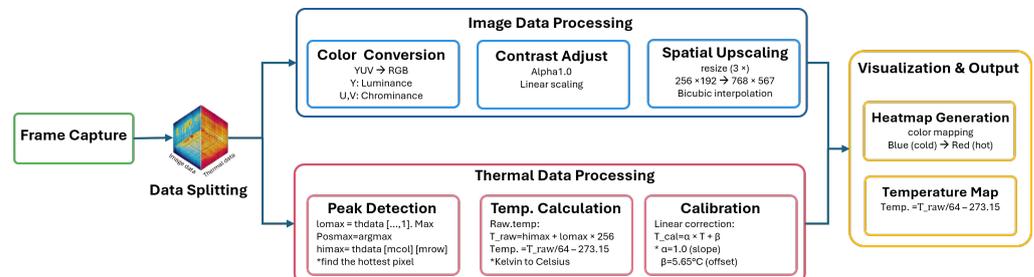
where  $H$  and  $W$  represent the height and width of a single modality, and

$$\mathbb{Z}_{256} = \{0, 1, \dots, 255\}$$

denotes the set of 8-bit integer values corresponding to raw pixel intensities. The raw frame is divided into two separate streams:

$$I_{\text{raw}} \xrightarrow{\text{split}} \left\{ I_{\text{img}} \in \mathbb{Z}_{256}^{H \times W}, D_{\text{th}} \in \mathbb{Z}_{256}^{H \times W \times 2} \right\}, \quad (2)$$

with  $I_{img}$  denoting the YUV image data and  $D_{th}$  denoting the paired bytes of thermal information. Each stream is subsequently processed within the same acquisition cycle but along independent pathways as presented in Figure 2: the  $I_{img}$  branch is converted into a color-mapped thermal image for visualization, whereas the  $D_{th}$  branch is decoded into pixel-wise temperature values, creating the quantitative temperature map corresponding to each thermal frame.



**Figure 2.** Processing pipeline of thermal camera acquisition. The raw frame is split into two streams: image data (top) for visualization and heatmap generation, and thermal data (bottom) for temperature calculation, calibration, and ROI detection. This split is specific to the TOPDON TC001 dual-stream frame format. The asterisk (\*) indicates that each block is further explained in the subsequent text to emphasize its specific role within the pipeline.

### 3.2.1. Image Data Processing

The image stream  $I_{img}$  undergoes a sequence of preprocessing operations to produce a heatmap suitable for visualization. The raw frames, initially captured in YUV format, are first converted into an RGB representation,

$$I_{RGB} = YUV2RGB(I_{img}), \quad (3)$$

the  $I_{RGB}$  denotes the color image obtained from  $I_{img}$ . To enhance visual clarity, a linear contrast adjustment is then applied, expressed as

$$I'_{RGB} = \alpha \times I_{RGB}, \quad \alpha = 1.0, \quad (4)$$

the  $\alpha$  denote as the contrast scaling factor. In this work,  $\alpha$  is fixed to unity, implying no additional scaling beyond the raw dynamic range.

Subsequently, the frame is spatially upscaled by bicubic interpolation:

$$I''_{RGB} = \text{Resize}(I'_{RGB}, W', H'), \quad W' = 3W, H' = 3H, \quad (5)$$

where  $\text{Resize}(\cdot)$  denotes the interpolation operator, and  $(W', H')$  are the target spatial dimensions set to three times the original resolution  $(W, H)$ . Finally, the enhanced frame is mapped into a false-color domain for visualization through

$$I_{heatmap} = \text{ColorMap}(I''_{RGB}), \quad (6)$$

the  $\text{ColorMap}(\cdot)$  maps the intensity distribution of  $I''_{RGB}$  into a perceptually enhanced heatmap representation for display visualization.

### 3.2.2. Thermal Data Processing

In the thermal data processing stage, pixel-wise temperatures are decoded from the paired bytes. Each thermal frame stores temperature values in two consecutive 8-bit values corresponding to the most significant byte (MSB) and least significant byte (LSB), which

are combined and linearly calibrated to form the quantitative temperature map shown in Figure 3. For each pixel coordinate  $(x, y)$  with

$$x \in \{0, \dots, W - 1\}, \quad y \in \{0, \dots, H - 1\}, \quad (7)$$

let  $H(x, y)$  and  $L(x, y)$  be the MSB and LSB, respectively, i.e.,  $H, L \in \mathbb{Z}_{256}$ . The raw temperature proxy is reconstructed as

$$R(x, y) = \frac{256 H(x, y) + L(x, y)}{64} - 273.15. \quad (8)$$

The combined value  $(256H(x, y) + L(x, y))$  represents raw temperature data encoded in Kelvin  $\times 64$  according to the manufacturer's format. Dividing by 64 converts this to Kelvin, and subtracting 273.15 yields the temperature in degrees Celsius, resulting in the calibrated temperature map  $R(x, y)$ . A linear calibration model is subsequently applied to compensate for sensor bias:

$$T(x, y) = \alpha R(x, y) + \beta, \quad (9)$$

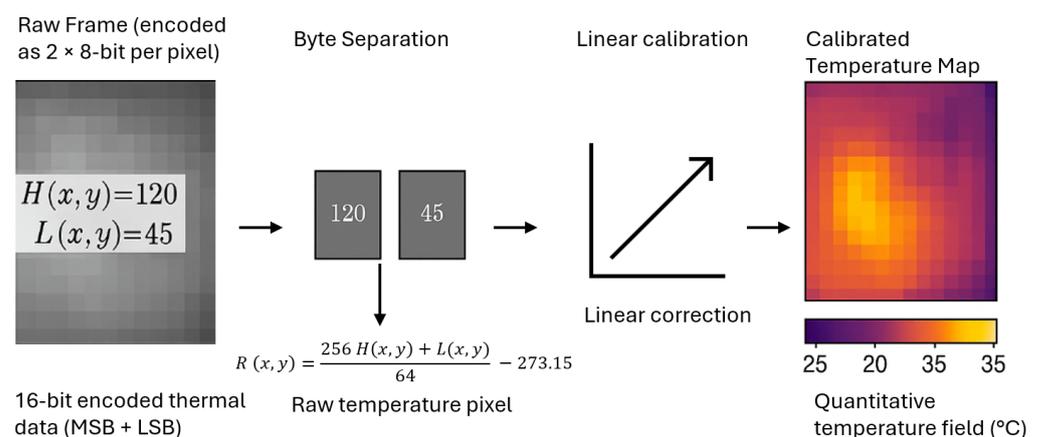
where  $\alpha$  is the calibration gain (scaling factor) and  $\beta$  is the calibration offset (bias in  $^{\circ}\text{C}$ ). The calibrated value  $T(x, y)$  corresponds to the corrected temperature at pixel  $(x, y)$ . And the maximum temperature within a frame is then localized as

$$(x^*, y^*) = \arg \max_{(x, y)} T(x, y), \quad T_{\max} = T(x^*, y^*), \quad (10)$$

where  $(x^*, y^*)$  indicates the pixel coordinates of the hottest point in the frame and  $T_{\max}$  is its corresponding temperature. For visualization, the calibrated temperature field is interpolated to generate a thermal map:

$$T_{\text{map}} = \text{Resize}(T(x, y), W', H'), \quad (11)$$

where  $\text{Resize}(\cdot, W', H')$  denotes a spatial interpolation operator mapping the original temperature matrix of size  $H \times W$  to a new resolution  $H' \times W'$  for display purposes. The resulting  $T_{\text{map}}$  provides the visualization branch of the processing pipeline.

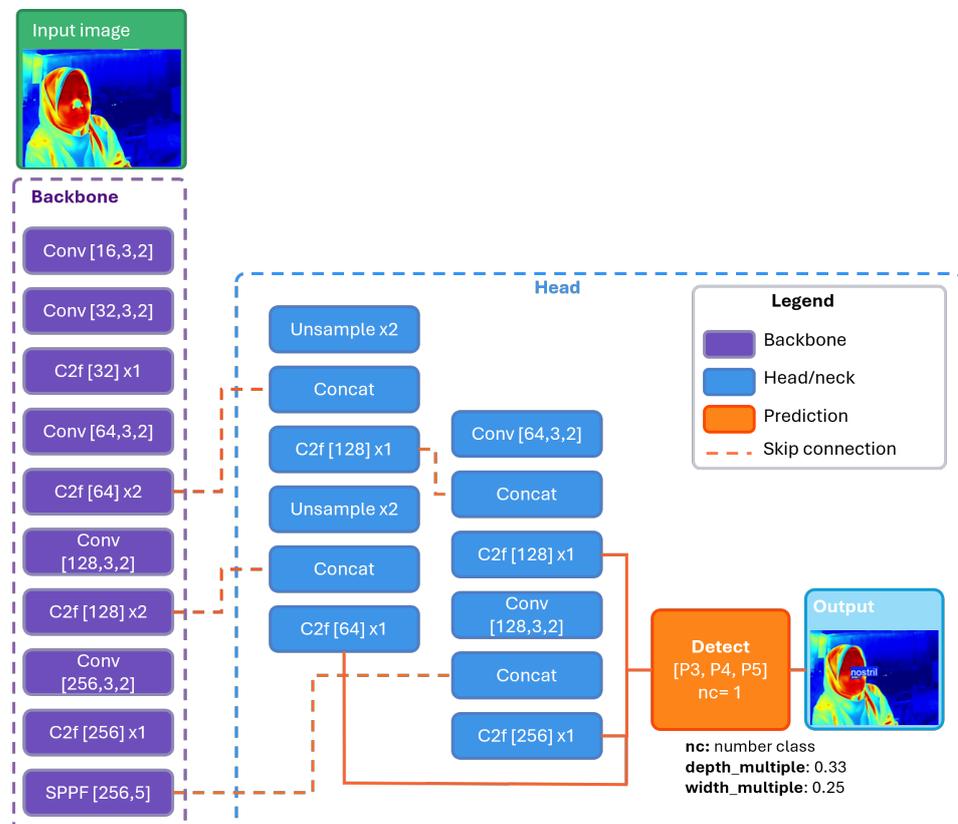


**Figure 3.** Thermal data decoding and calibration process. Each pixel is formed from two 8-bit bytes (MSB + LSB) combined to reconstruct  $R(x, y)$ , followed by linear calibration to produce the quantitative temperature map ( $^{\circ}\text{C}$ ).

### 3.3. ROI Localization and Temperature Feature Extraction

#### 3.3.1. YOLO-Based Nostril Detection

Respiratory monitoring based on wearable sensors is often limited by discomfort, movement artifacts, and the need for frequent recalibration [44,45]. RGB video methods offer a non-contact alternative but remain highly sensitive to illumination changes, motion artifacts, and computational overhead [46–48]. Thermal imaging provides a more suitable modality because it is independent of lighting conditions and relatively robust to minor head movements, making it advantageous for continuous monitoring. Yet the low resolution and limited texture of thermal frames make nostril localization challenging. To overcome this difficulty, a YOLO-based detection model was adopted, as illustrated in Figure 4. YOLO was chosen for its balance between detection accuracy and computational efficiency, which makes it suitable for real-time deployment on embedded edge hardware. Unlike RGB data, thermal frames contain only coarse temperature gradients and lack color cues, which reduces the effectiveness of standard feature extraction; accordingly, architectural and training adaptations were required.



**Figure 4.** Architecture of the YOLO-based nostril detection model for thermal imagery. The backbone extracts multi-scale thermal features, while the head performs feature fusion and detection across P3–P5 layers to localize the nostril region in real time. The blue dashed outline indicates the boundary of the detection head, and overlapping elements are used for compact visualization without affecting architectural interpretation.

In this work, YOLOv8n was selected as the object detection framework for its computational efficiency and suitability to real-time thermal imaging. The lightweight model architecture enables effective feature extraction while maintaining low computational cost, which is appropriate for the relatively simple thermal domain where nostril regions are primarily defined by local hot-cold gradients rather than complex textures. The detection head maintains high-resolution feature maps, improving the sensitivity of the model to small ROI that occupy less than five percent of the thermal frame. In parallel, the training

procedure was tailored to the thermal modality. A dataset of 7958 annotated thermal images (7113 training, 563 validation, 282 testing) was assembled, incorporating variations in head orientation, distance, and partial occlusion. Data augmentation strategies avoided color-based transformations and instead emphasized brightness and contrast adjustments, Gaussian noise injection, and mild geometric perturbations to reproduce sensor variability and natural subject motion.

For deployment, the trained detector is executed using the Ultralytics YOLO runtime on the embedded device, without reliance on external deep-learning frameworks or GPU acceleration. To reduce computational load, detection is performed on every even frame, and a Kalman filter predicts the ROI on intermediate frames. With frame index  $k \in \mathbb{Z}_{\geq 0}$  and camera rate  $f_{\text{cam}}$  (Hz), the detection schedule is

$$\mathcal{S}_{\text{det}} = \{k \mid k \bmod 2 = 0\} \quad (\text{stride } s = 2), \quad (12)$$

which yields the effective detection rate

$$f_{\text{det}} = \frac{f_{\text{cam}}}{s} = \frac{f_{\text{cam}}}{2}. \quad (13)$$

The ROI used at frame  $k$  is obtained from the detector when  $k \in \mathcal{S}_{\text{det}}$  (and a detection exists) and from the Kalman prediction otherwise:

$$\text{ROI}_k = \begin{cases} \text{ROI}_k^{\text{det}}, & k \in \mathcal{S}_{\text{det}} \text{ and a detection exists,} \\ \text{ROI}_k^{\text{trk}}, & \text{otherwise.} \end{cases} \quad (14)$$

### 3.3.2. Kalman Filter Tracking

Since YOLO-based generates nostril detections only on even frames, the Kalman filter predicts the ROI on odd frames and whenever an even-frame detection is unavailable. To provide stable localization of the nostril region, an eight-dimensional Kalman filter jointly estimates the bounding-box position and its temporal dynamics. The state vector is defined as

$$\mathbf{x}_k = [c_x, c_y, w, h, \dot{c}_x, \dot{c}_y, \dot{w}, \dot{h}]^T, \quad (15)$$

where  $(c_x, c_y)$  are the bounding-box center coordinates (pixels),  $(w, h)$  its width and height (pixels), and the dotted variables the corresponding temporal velocities.

The temporal evolution of the state follows a discrete constant-velocity model:

$$\mathbf{x}_k = \mathbf{F} \mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad \mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad (16)$$

where  $\mathbf{F}$  is the  $8 \times 8$  transition matrix,  $\mathbf{w}_{k-1}$  the process noise, and  $\mathbf{Q}$  its covariance. Using the frame interval  $\Delta t = 1/f_{\text{cam}}$ , the transition matrix is

$$\mathbf{F} = \begin{bmatrix} \mathbf{I}_4 & \Delta t \mathbf{I}_4 \\ \mathbf{0}_4 & \mathbf{I}_4 \end{bmatrix}, \quad (17)$$

with  $\mathbf{I}_4$  the  $4 \times 4$  identity matrix and  $\mathbf{0}_4$  the  $4 \times 4$  zero matrix. The choice of the constant-velocity state-space model in Equations (15) and (16) follows standard formulations widely used in visual object tracking, as it provides a minimal yet sufficiently expressive representation of bounding-box motion [49,50]. The transition matrix  $\mathbf{F}$  is block-structured with identity and  $\Delta t \mathbf{I}_4$  submatrices, which yields eigenvalues equal to 1. Consequently, the discrete-time system is marginally stable, as expected for constant-velocity motion models; the continuous-time Hurwitz condition does not directly apply in this setting. Since the

Kalman filter is employed solely for state estimation rather than control, controllability is not required. The pair  $(\mathbf{F}, \mathbf{H})$  is observable, as the associated observability matrix has full rank for any  $\Delta t > 0$ , ensuring that all components of the state vector, including position, size, and their velocities are inferable from the detector measurements.

At each frame, the YOLO-based detector produces a bounding box with corners  $(x_1, y_1)$  (top-left) and  $(x_2, y_2)$  (bottom-right). This is converted into the measurement vector

$$\mathbf{z}_k = \left[ \frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}, x_2 - x_1, y_2 - y_1 \right]^\top, \quad (18)$$

which contains the observed center position and box dimensions. The measurement model is

$$\mathbf{z}_k = \mathbf{H} \mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad \mathbf{H} = [\mathbf{I}_4 \ \mathbf{0}_{4 \times 4}]. \quad (19)$$

Let  $d_k \in \{0, 1\}$  indicate whether a detector output exists at frame  $k$  (on scheduled even frames). The measurement-use indicator is

$$m_k = \begin{cases} 1, & \text{if } k \bmod 2 = 0 \text{ and } d_k = 1, \\ 0, & \text{otherwise,} \end{cases} \quad (20)$$

and a fixed measurement covariance is used

$$\mathbf{R}_k = \mathbf{R}. \quad (21)$$

The recursion runs on every frame. Prediction:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F} \hat{\mathbf{x}}_{k-1}, \quad \mathbf{P}_{k|k-1} = \mathbf{F} \mathbf{P}_{k-1} \mathbf{F}^\top + \mathbf{Q}. \quad (22)$$

Update (only if  $m_k = 1$ ):

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{H} \hat{\mathbf{x}}_{k|k-1}, \quad \mathbf{S}_k = \mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^\top + \mathbf{R}_k, \quad (23)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}^\top \mathbf{S}_k^{-1}, \quad \hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \mathbf{y}_k, \quad (24)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_{k|k-1}. \quad (25)$$

If  $m_k = 0$ , the prediction becomes the current estimate. The output box is reconstructed as

$$x_1 = c_x - \frac{w}{2}, \quad y_1 = c_y - \frac{h}{2}, \quad x_2 = c_x + \frac{w}{2}, \quad y_2 = c_y + \frac{h}{2}. \quad (26)$$

This even-odd schedule halves detector invocations, provides ROI estimates for skipped frames via Kalman prediction, and preserves temporal continuity under short dropouts, head motion, or partial occlusion.

### 3.3.3. Temperature Extraction

Once the nostril ROI is localized by the detection-tracking pipeline, the calibrated thermal image at frame  $k$  is treated as a discrete grid  $T_k[x, y]$  (temperature in  $^\circ\text{C}$  at pixel  $(x, y)$ ). The ROI is the integer-indexed set

$$\text{ROI}_k = \{(x, y) \in \mathbb{Z}^2 \mid x_1^{(k)} \leq x \leq x_2^{(k)}, y_1^{(k)} \leq y \leq y_2^{(k)}\}, \quad (27)$$

where  $(x_1^{(k)}, y_1^{(k)})$  and  $(x_2^{(k)}, y_2^{(k)})$  are the top-left and bottom-right corners of the bounding box at frame  $k$ . The representative nostril temperature for frame  $k$  is the minimum within the ROI,

$$\hat{T}_k = \min_{(x,y) \in \text{ROI}_k} T_k[x, y], \quad (28)$$

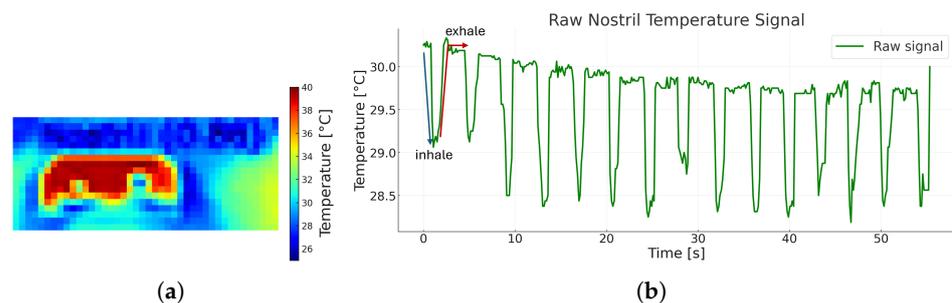
and the pixel attaining this minimum is recorded as

$$p_k^* = \arg \min_{(x,y) \in \text{ROI}_k} T_k[x, y]. \quad (29)$$

Selecting the coldest pixel yields a physiologically consistent proxy of airflow, as inhalation introduces cooler ambient air whereas exhalation releases warmer expired air. Across frames, the extracted temperatures form the scalar sequence

$$\mathcal{S} = \{\hat{T}_k\}_{k=0}^{N-1}, \quad (30)$$

where  $N$  is the number of samples retained in the observation buffer. A sliding buffer of approximately 20 s (i.e.,  $N \approx \lceil 20 f_{\text{cam}} \rceil$ ) preserves multiple respiratory cycles while maintaining responsiveness for real-time monitoring. This raw nostril-temperature signal is then used for band-pass filtering, phase detection, and respiratory-rate estimation. Figure 5 illustrates the extraction result: panel (a) shows the cropped thermal ROI with the coldest-temperature pixel at each frame, and panel (b) shows the resulting raw sequence  $\mathcal{S}$ , whose cyclic oscillations align with inhalation (cooling) and exhalation (warming) events.



**Figure 5.** Illustration of respiratory signal extraction. (a) Cropped nostril ROI represented in the thermal domain; (b) corresponding raw nostril temperature signal  $\mathcal{S}$ , where the oscillations of  $\hat{T}_k$  denote inhalation (cooling) and exhalation (warming), indicated by the blue and red lines, respectively, providing the input for subsequent signal processing and respiratory rate estimation.

### 3.4. Adaptive Breathing Phase Detection and Respiratory Rate Calculation

#### 3.4.1. Adaptive Breathing Phase Detection

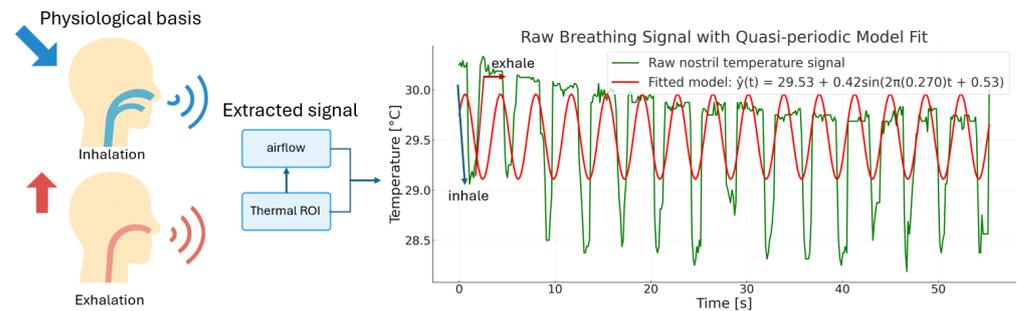
The breathing phase pipeline starts from the illustration in Figure 6: inhalation cools the nostril surface, whereas exhalation warms it. After ROI extraction, each frame  $k$  provides a scalar sample  $\hat{T}[k]$  (in  $^{\circ}\text{C}$ ). With camera frame rate  $f_{\text{cam}}$  (Hz), samples occur at  $t_k = k/f_{\text{cam}}$ ; equivalently,

$$\hat{T}[k] = \hat{T}(t_k) + \varepsilon_k, \quad (31)$$

where  $\varepsilon_k$  denotes measurement noise. The processing that follows operates on the discrete sequence  $\{\hat{T}[k]\}$ . The raw sequence shows alternating cooling and warming, but is also affected by drift and noise. To describe its expected structure and guide filter design, it is convenient to write the quasi-periodic model

$$\hat{T}(t) = T_0 + A \sin(2\pi f_{\text{RR}} t + \phi) + d_{\text{low}}(t) + n_{\text{high}}(t), \quad (32)$$

where  $T_0$  is the baseline temperature,  $A$  the oscillation amplitude,  $f_{RR}$  the respiratory frequency (Hz),  $\phi$  the phase,  $d_{low}(t)$  a slow drift term, and  $n_{high}(t)$  high-frequency noise. This model is illustrative; the digital operations below use the sampled signal  $\hat{T}[k]$ .



**Figure 6.** Schematic illustration of the proposed respiration monitoring framework. Inhalation of cooler air lowers nostril temperature, while exhalation of warmer air raises it (**left**), illustrated by blue and red arrows, respectively. The airflow-induced modulation is captured from the thermal ROI and converted into a raw nostril temperature signal (**middle**). The extracted signal (green) exhibits quasi-periodic fluctuations that can be approximated by a sinusoidal model (red), reflecting the underlying breathing cycles (**right**).

Since respiration lies in the 0.08–0.7 Hz band, a 4th-order Butterworth band-pass filter with cutoffs at 0.08 and 0.7 Hz is applied. The filter is implemented with forward–backward recursion to ensure a zero-phase response, thereby preserving the temporal integrity of breathing cycles. The band-pass output satisfies the standard IIR difference equation:

$$\hat{T}_{bp}[n] = \sum_{m=0}^M b_m \hat{T}[n-m] - \sum_{j=1}^N a_j \hat{T}_{bp}[n-j], \quad (33)$$

where  $b_m$  and  $a_j$  are the Butterworth coefficients. This passband covers 5–42 BPM and suppresses baseline drift and high-frequency noise, preserving breath timing. Building on the zero-phase band-passed output  $\hat{T}_{bp}[n]$ , the local heating/cooling trend is quantified using a rectangular moving-average window of length  $W$ . The discrete velocity surrogate is defined as the difference between two adjacent moving averages:

$$v[n] = (r_W * \hat{T}_{bp})[n] - (r_W * \hat{T}_{bp})[n-W], \quad (34)$$

where “\*” denotes discrete convolution and  $r_W[k]$  is the rectangular (uniform) kernel

$$r_W[k] = \begin{cases} \frac{1}{W}, & 0 \leq k \leq W-1, \\ 0, & \text{otherwise.} \end{cases} \quad (35)$$

An equivalent summation form is

$$v[n] = \frac{1}{W} \sum_{k=0}^{W-1} \hat{T}_{bp}[n-k] - \frac{1}{W} \sum_{k=W}^{2W-1} \hat{T}_{bp}[n-k], \quad (36)$$

under the zero-phase design, the sign of  $v[n]$  aligns with physiology:

$$v[n] > 0 \Rightarrow \text{exhalation (warming)}, \quad v[n] < 0 \Rightarrow \text{inhalation (cooling)}, \quad (37)$$

in implementation, a small window (e.g.,  $W = 3$  samples) attenuates frame-to-frame jitter while preserving phase timing within the 5–42 BPM operating range.

To normalize across subjects and amplitudes, an adaptive, data-dependent threshold is derived from the most recent  $L$  velocity samples. Let the length- $L$  window be

$$V_n \triangleq \{v[n-i] \mid i = 0, 1, \dots, L-1\}. \quad (38)$$

The location statistic and dispersion are defined by the sample median and the median absolute deviation (MAD):

$$m_v[n] = \text{median}(V_n), \quad (39)$$

$$\text{MAD}_v[n] = \text{median}_{i \in \{0, \dots, L-1\}} |v[n-i] - m_v[n]|. \quad (40)$$

A symmetric, scale-invariant threshold is then

$$\theta[n] = \alpha \text{MAD}_v[n] + \varepsilon, \quad (41)$$

with sensitivity coefficient  $\alpha > 0$  (e.g.,  $\alpha = 0.6$ ), a short history length  $L$  (e.g., 15–25 samples), and a small  $\varepsilon > 0$  to avoid degeneracy when variability is minimal. The threshold is applied symmetrically around zero to map velocity to the physiological phase:

$$v[n] \geq \theta[n] \Rightarrow \text{exhalation (warming)}, \quad v[n] \leq -\theta[n] \Rightarrow \text{inhalation (cooling)}. \quad (42)$$

Optionally, a minimum dwell converts threshold crossings into stable segments; with sampling frequency  $f_{\text{cam}}$ , the dwell length in samples is

$$N_{\text{min}} = \lceil t_{\text{min}} f_{\text{cam}} \rceil, \quad (43)$$

ensuring physiologically plausible durations within the 5–42 BPM operating range.

To convert the thresholded velocity into stable phase labels, a hysteretic state machine with a minimum dwell is employed. Let the instantaneous phase label be

$$s[n] \in \{-1, 0, +1\}, \quad (44)$$

representing Inhalation ( $-1$ ), Neutral/Hold ( $0$ ), and Exhalation ( $+1$ ). The dwell requirement expressed in samples Equation (43), ensuring physiologically plausible segment durations within the 5–42 BPM operating range. Transitions are gated by the elapsed persistence of the current state. Let  $\tau[n-1]$  denote the number of consecutive samples that state  $s[n-1]$  has persisted up to time  $n-1$ . The phase update is

$$s[n] = \begin{cases} +1, & \text{if } v[n] \geq \theta[n] \wedge \tau[n-1] \geq N_{\text{min}}, \\ -1, & \text{if } v[n] \leq -\theta[n] \wedge \tau[n-1] \geq N_{\text{min}}, \\ s[n-1], & \text{otherwise.} \end{cases} \quad (45)$$

The persistence counter is updated recursively as

$$\tau[n] = \begin{cases} 0, & \text{if } s[n] \neq s[n-1], \\ \tau[n-1] + 1, & \text{if } s[n] = s[n-1]. \end{cases} \quad (46)$$

Optionally, brief near-threshold fluctuations may be represented as a neutral state to emphasize ambiguity around zero velocity:

$$(\text{optional}) \quad s[n] \leftarrow 0 \quad \text{if } |v[n]| < \theta[n]. \quad (47)$$

This hysteretic formulation suppresses chatter from transient perturbations while preserving accurate timing of inhalation and exhalation transitions.

Following the hysteretic phase labeling, brief flicker patterns of the form  $A \rightarrow B \rightarrow A$  are suppressed by merging the short middle segment into its flanking phase. Here  $A, B \in \{-1, 0, +1\}$  denote the phase labels in Equation (44). Let  $\{n_i\}_{i \geq 0}$  denote the ordered change-points of  $s[n]$ ,

$$n_0 = 0, \quad n_{i+1} = \min\{n > n_i \mid s[n] \neq s[n-1]\}, \quad (48)$$

and define the  $i$ -th segment state and duration by

$$q_i \triangleq s[n] \text{ for all } n \in [n_i, n_{i+1} - 1], \quad L_i \triangleq n_{i+1} - n_i. \quad (49)$$

With sampling frequency  $f_{\text{cam}}$ , the sample-based consolidation threshold is

$$N_c = \lceil \tau_c f_{\text{cam}} \rceil, \quad (50)$$

where  $\tau_c$  is a short duration (e.g.,  $\tau_c = 0.3$  s) chosen to reject physiologically implausible micro-segments within 5–42 BPM.

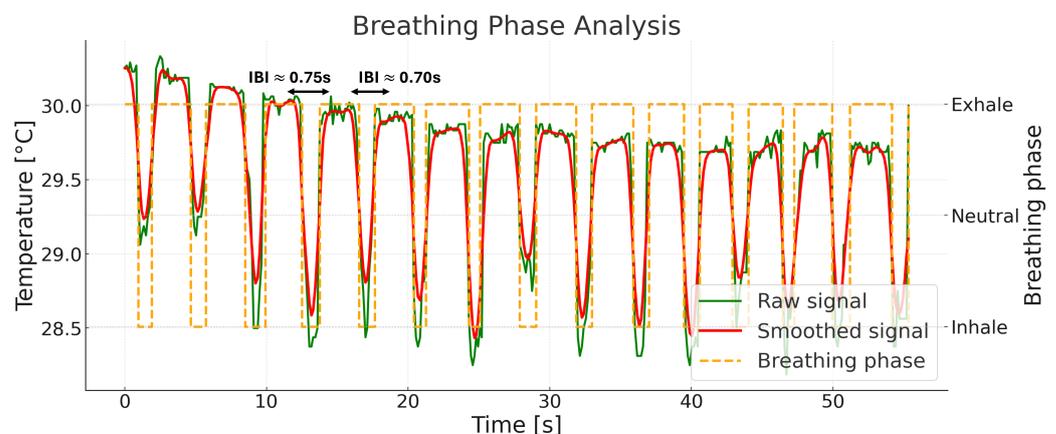
The consolidation rule replaces the short intermediate phase  $B$  (i.e.,  $q_i$ ) by its flanking phase  $A$  (i.e.,  $q_{i-1} = q_{i+1}$ ) whenever an  $A \rightarrow B \rightarrow A$  pattern occurs and the intermediate duration is below  $N_c$ :

$$\text{if } q_{i-1} = q_{i+1} \wedge q_i \neq q_{i-1} \wedge L_i < N_c, \quad \text{then } s[n] \leftarrow q_{i-1} \quad \forall n \in [n_i, n_{i+1} - 1]. \quad (51)$$

This procedure can be applied iteratively over  $\{(q_{i-1}, q_i, q_{i+1})\}$  until no violations remain, yielding a piecewise-constant phase trace without short-lived toggles and enabling stable inter-breath-interval and respiratory-rate estimation.

### 3.4.2. Respiratory Rate Calculation

Respiratory rate is derived from the consolidated phase labels  $s[n]$  described previously. Figure 7 illustrates the outputs used here: the raw nostril-temperature sequence (green), its smoothed version (red), and the phase trace (orange) that marks inhalation, neutral/hold, and exhalation segments. Inter-breath intervals (IBI) are computed from consecutive phase transitions of a chosen event type (exhalation onsets in this implementation).



**Figure 7.** Breathing phase detection from nostril temperature signals. Raw (green) and smoothed (red) sequences are shown, with detected breathing phases indicated by orange dashed lines corresponding to exhalation phases; inhalation and hold are implicitly identified by phase transitions.

Let  $f_{\text{cam}}$  be the camera frame rate and define the exhalation-onset event set

$$\mathcal{E} \triangleq \{ k \mid s[k-1] \neq +1 \wedge s[k] = +1 \}. \quad (52)$$

For consecutive events  $k_{i-1}, k_i \in \mathcal{E}$  with  $k_i > k_{i-1}$ , the IBI (seconds) is

$$\Delta t_i = \frac{k_i - k_{i-1}}{f_{\text{cam}}}. \quad (53)$$

Physiological plausibility is enforced consistently with the 5–42 BPM operating band by accepting only

$$\Delta t_{\min} \leq \Delta t_i \leq \Delta t_{\max}, \quad \Delta t_{\min} = \frac{60}{42} \text{ s}, \quad \Delta t_{\max} = \frac{60}{5} \text{ s}. \quad (54)$$

Each validated interval yields an instantaneous respiratory rate

$$RR_i = \frac{60}{\Delta t_i} \text{ (BPM)}. \quad (55)$$

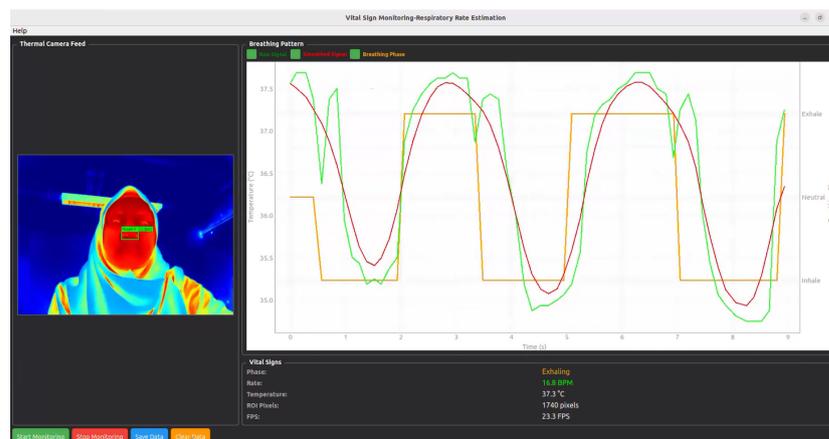
To obtain a stable yet responsive trace as in Figure 7, two lightweight smoothers are applied sequentially: a causal weighted update

$$\widetilde{RR}_i = 0.6 RR_i + 0.4 \widetilde{RR}_{i-1}, \quad \widetilde{RR}_0 = RR_0, \quad (56)$$

followed by an exponential moving average (EMA) with coefficient  $\alpha = 0.7$ ,

$$RR_{\text{final},i} = \alpha \widetilde{RR}_i + (1 - \alpha) RR_{\text{final},i-1}, \quad RR_{\text{final},0} = \widetilde{RR}_0. \quad (57)$$

This IBI  $\rightarrow$  weighted-average  $\rightarrow$  EMA pipeline yields a robust BPM estimate under motion, thermal drift, and noise while remaining suitable for real-time embedded execution. For real-time visualization, a PyQt5 window renders the thermal video feed with ROI detection result overlays alongside the breathing waveform panel as well as the current breathing phase, respiratory rate in BPM, nostril temperature, ROI pixel area and the FPS of system as illustrates in Figure 8.



**Figure 8.** Graphical user interface (GUI) of the proposed respiratory rate monitoring system. The left panel shows the thermal camera feed with automated nostril detection. The right panel illustrates the extracted nostril temperature signal, the green, red, and orange lines denote the raw signal, smoothed signal, and detected breathing phase, respectively. The bottom panel provides real-time vital signs, including current breathing phase, respiratory rate in BPM, nostril temperature, and processing frame rate (FPS).

## 4. Experimental Results

### 4.1. Hardware and Software Configuration

The respiratory rate monitoring system was implemented on a Jetson Orin Nano Developer Kit (NVIDIA Corporation, Santa Clara, CA, USA; 6-core ARM Cortex-A78AE CPU, 8 GB LPDDR5 RAM) running Ubuntu 20.04 LTS. The YOLO-based nostril detection and respiratory-signal processing algorithms were developed in Python 3.10 with OpenCV 4.x and executed directly on the embedded device. Thermal video streams were processed in real-time, with inference and signal analysis performed entirely on the edge device, without requiring cloud-based computation. The raw temperature data were continuously collected from the TOPDON TC001 thermal camera (Topdon Technology Co., Ltd., Shenzhen, China) with a resolution of  $256 \times 192$  pixels and a lightweight design (30 g). The camera was connected directly to the Jetson Orin Nano for on-device processing, ensuring minimal latency and maintaining a non-intrusive measurement environment.

To quantify the computational complexity of the proposed system on embedded hardware, the end-to-end execution was profiled over a 60 s continuous run on the Jetson Orin Nano. The average per-frame processing time was 65.2 ms, with a 95th-percentile latency of 85.3 ms, indicating that 95% of frames were processed within this bound. This corresponds to a real-time throughput of 22.5 FPS, with a standard deviation of 1.8 FPS, reflecting stable runtime characteristics throughout the measurement interval. As shown in Table 1, YOLO-based nostril detection constituted the primary computational load (35.5 ms, 54.5%), followed by thermal capture latency (15.0 ms, 23.3%) and graph/GUI updates (10.8 ms, 16.6%). All remaining modules, including temperature extraction (2.8 ms), adaptive phase detection and signal processing (1.5 ms), Kalman tracking (1.2 ms), and IBI calculation (0.5 ms), each contributed less than 5% of the total per-frame cost. System resource monitoring further showed moderate utilization, with mean CPU usage of 42.5%, GPU usage of 68.2%, and memory consumption of 850 MB. These results confirm that the computational burden is lightweight and well within the real-time operating envelope of low-power embedded edge devices.

**Table 1.** Profiling summary of computational performance on the Jetson Orin Nano (60 s run).

<b>(A) Runtime Breakdown per Component</b>					
Component	Mean (ms)	p95 (ms)	Percentage		
Total Frame Time	65.20	85.30	100%		
YOLO Detection	35.50	42.10	54.5%		
Thermal Capture	15.00	30.00	23.3%		
Graph/GUI Update	10.80	14.20	16.6%		
Signal Processing	3.50	3.80	5.4%		
Temperature Extraction	2.80	3.50	4.3%		
Kalman Tracking	1.20	0.80	1.8%		
IBI Calculation	0.50	0.80	0.8%		
<b>(B) System Resource Utilization</b>			<b>(C) End-to-End Throughput Statistics</b>		
Metric	Mean	p05	p95	Metric	Value (FPS)
CPU Usage	42.5%	35.2%	58.3%	Mean FPS	22.5
GPU Usage	68.2%	67.5%	80.3%	Min FPS	18.2
Memory Usage	850.3 MB	820.5 MB	890.2 MB	Max FPS	28.8
				p50	24.5
				p95	24.5
				Std Dev	1.8

#### 4.2. Respiratory Rate Experimental Procedures

A pilot study was conducted with ten healthy adults ( $N = 10$ , aged  $33.3 \pm 4.38$ ) under institutional ethics approval and written informed consent. Ground-truth  $RR$  was obtained by dual-rater manual counts from the recordings; for metronome-paced blocks, the target rate was logged as an auxiliary reference. To establish ground-truth  $RR$ , manual tally counting [51] was performed on the experimental video recordings. Each breathing cycle was visually identified by observing airflow from the nostril, and the total number of cycles within a predefined time window was recorded. The respiratory rate ( $RR$ ) was then calculated as:

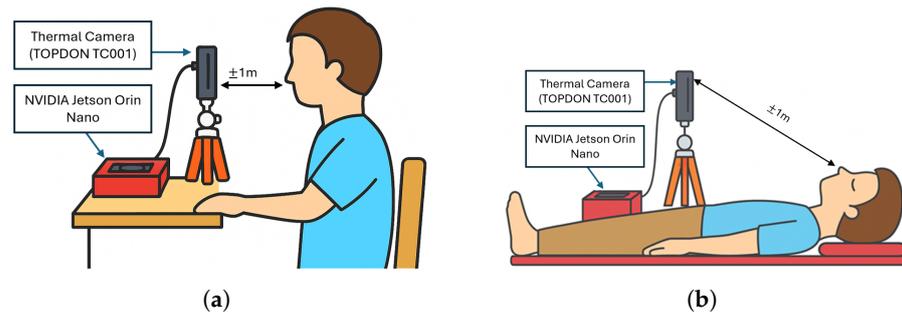
$$RR = \frac{N_{\text{breaths}}}{T} \times 60 \quad [\text{bpm}], \quad (58)$$

where  $N_{\text{breaths}}$  denotes the number of observed breathing cycles and  $T$  represents the duration of the observation in seconds. For example, if 22 breaths were observed during a 60-s video, the reference  $RR$  was 22 bpm.

The experimental protocol, summarized in Table 2 and the subject faced the thermal camera as illustrated in Figure 9, was designed not only to validate respiratory rate ( $RR$ ) estimation under controlled conditions but also to emulate scenarios relevant to long-lie incidents. In a long-lie situation, an individual may remain immobile for an extended duration in various postures or under partial occlusions, where reliable respiration monitoring becomes a key indicator of consciousness and vitality. The resting and paced breathing sessions establish baseline accuracy across normal and rhythmic respiration patterns, forming the reference for physiological consistency. The robustness (soft speech) condition introduces mild facial motion to evaluate tolerance to articulation, representative of irregular speech or groaning that may occur before or after a fall. The distance and off-axis yaw conditions simulate variations in camera placement and subject orientation that would naturally arise when the person is lying at different angles or when the thermal sensor is mounted in a fixed overhead position. Finally, the posture (supine) recordings directly mimic a post-fall scenario, where the subject lies facing upward with minimal motion. Collectively, these conditions ensure that the proposed system is trained and validated under realistic variability, facilitating robust respiratory monitoring during long-lie detection.

**Table 2.** Experimental design overview for respiratory rate data collection. Each block is a 60 s recording.

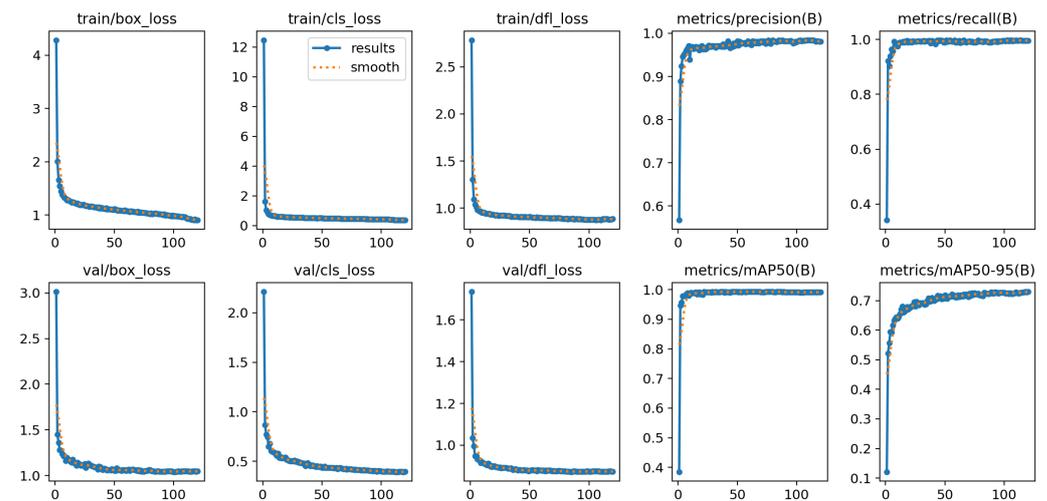
Condition Set	Blocks per Subject	Duration	Description
Resting (spontaneous)	2×	60 s	Seated, natural nasal breathing, mouth closed
Paced breathing (metronome)	3×	60 s	Seated; guided at 12, 18, 24 BPM (randomized order); metronome target logged as auxiliary reference
Robustness (soft speech)	2×	60 s	Seated; counting aloud to emulate mild articulatory motion
Distance (stood)	3×	60 s	Spontaneous breathing at 1.0, 1.5, and 2.0 m; camera height/pitch held constant
Off-axis yaw (seated)	2×	60 s	Spontaneous breathing at $\pm 30^\circ$ yaw; neutral pitch/roll instructed
Posture (supine)	2×	60 s	Spontaneous breathing in supine facing camera; camera pitched downward



**Figure 9.** Illustration of experimental setup. (a) Seated configuration: thermal camera aligned at nose height at 1–2 m to capture nostril temperature; (b) Supine configuration: camera pitched downward to preserve nasal visibility.

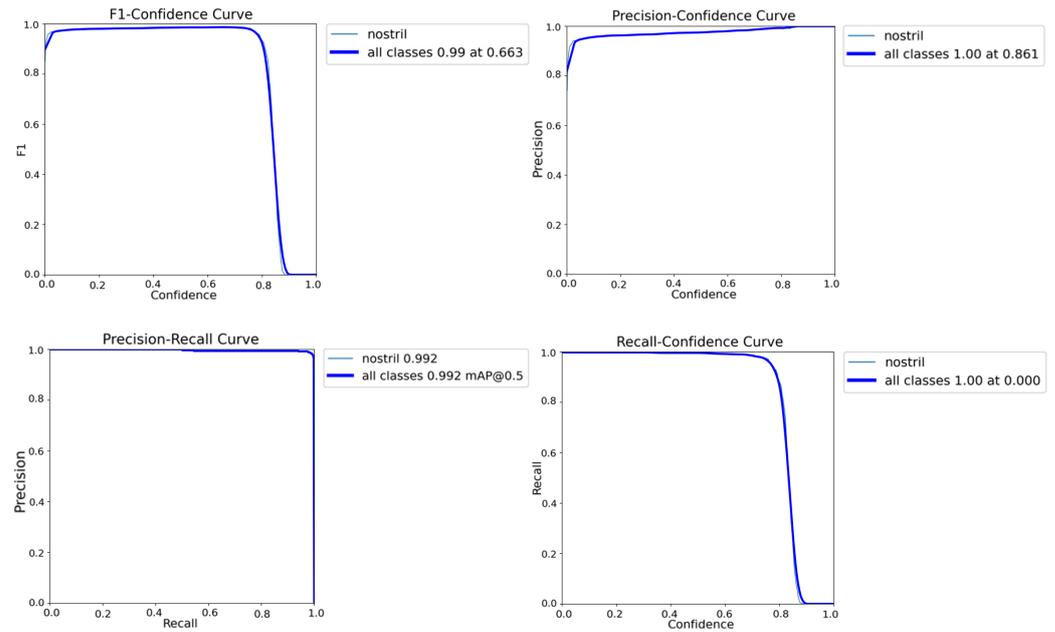
#### 4.3. Nostril Detection Performance

The nostril detection model was trained using a YOLO-based architecture, showing rapid and stable convergence, as evidenced by the steady decrease in box, classification, and distribution focal losses across both the training and validation sets, as illustrated in Figure 10. The evaluation metrics, including precision, recall, mAP@0.5, and mAP@0.5–0.95, exhibit consistent improvement and stabilization across epochs, confirming the robustness and generalization capability of the trained model for reliable nostril localization in thermal imagery.



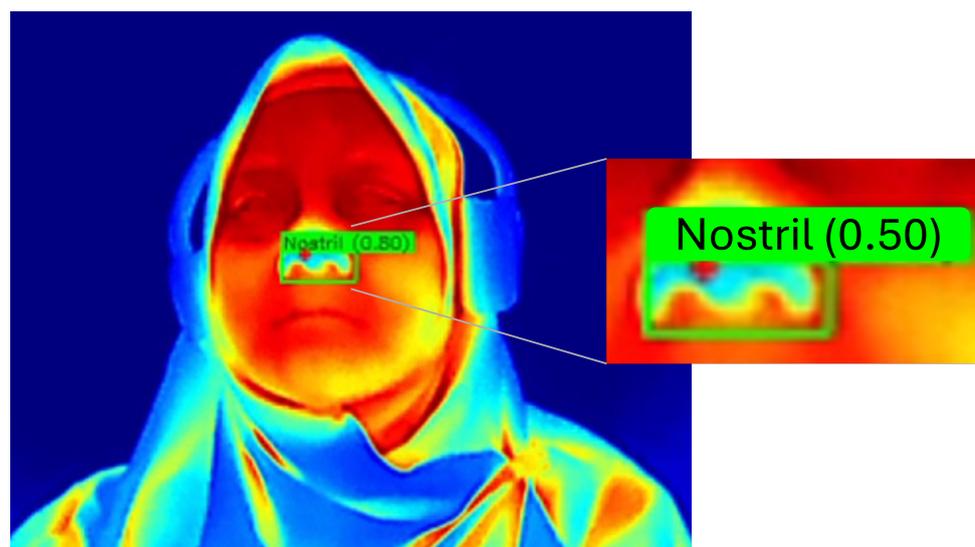
**Figure 10.** Training performance curves of the YOLO-based nostril detection model. The plots show the evolution of box loss, classification loss, distribution focal loss (DFL), and evaluation metrics (precision, recall, mAP@0.5, and mAP@0.5–0.95) for both training and validation datasets over the training epochs.

The training curves demonstrate a steady increase in precision and recall, reaching over 99% within the first few epochs. Quantitative evaluation metrics in Figure 11 further confirm the model's robustness, with the Precision–Recall (PR) curve showing an area under the curve (AUC) of 0.992, and the F1–Confidence curve peaking at 0.99. Both Precision–Confidence and Recall–Confidence curves indicate stable predictions across a wide confidence range, with optimal performance observed at a confidence threshold of approximately 0.86.

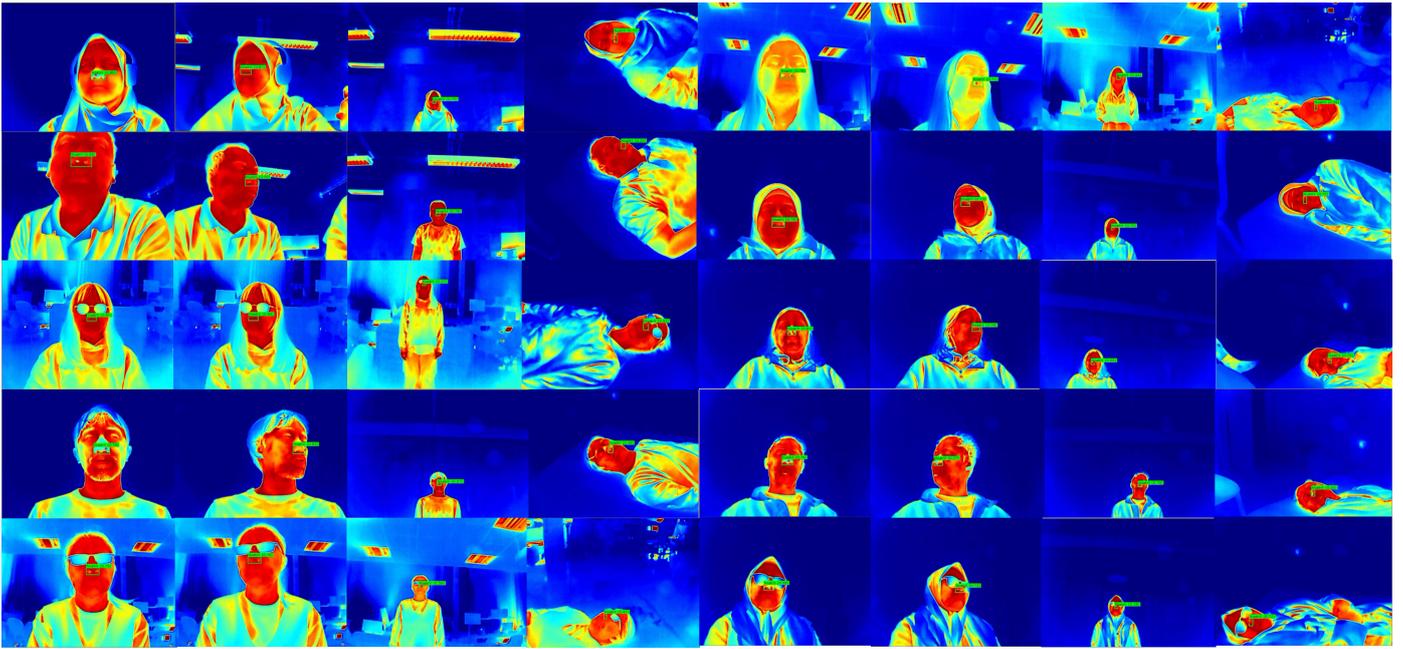


**Figure 11.** Performance curves of the YOLO-based nostril detection model on the validation dataset including F1–Confidence curve, Precision–Confidence curve, Precision–Recall (PR) curve, and Recall–Confidence curve. The results indicate consistently high precision, recall, and F1-scores across a wide confidence range, with optimal detection performance achieved at confidence thresholds between 0.66 and 0.86.

To complement the quantitative analysis, Figure 12 provides an enlarged view of the nostril detection output, making the detection label and bounding box clearly visible. Meanwhile, Figure 13 presents qualitative examples of the YOLO-based nostril detector applied to diverse thermal video frames. These examples demonstrate consistent and reliable nostril localization under various conditions, including different head poses, lighting variations, and partial occlusions. The trained model accurately identifies the nostril region across diverse thermal video frames, demonstrating robustness and stability for downstream respiratory rate estimation tasks.



**Figure 12.** Close-up example of the detected nostril region. The enlarged crop highlights the detection label and bounding box that are less visible in the full-resolution montage presented in Figure 13.



**Figure 13.** Qualitative results of the YOLO-based nostril detection model on thermal video frames. The model consistently localizes the nostril region with high confidence across varying head poses, facial orientations, and partial occlusions.

#### 4.4. Respiratory Rate Estimation Accuracy

Before estimating the respiratory rate following the procedures in Section 4.2, the nostril detector was first validated to ensure reliable localization across different subject poses, as illustrated in Figure 14, showing the thermal frame was collected and aligned with the experimental setup in Table 2. The automatically detected nostril regions (green bounding boxes) are shown across a wide range of conditions, including resting, metronome-paced breathing, soft-speech influence, distance variation, off-axis head orientation, and posture changes. This variability ensures that the evaluation reflects realistic operating scenarios with differences in viewpoint, articulation, and body orientation.

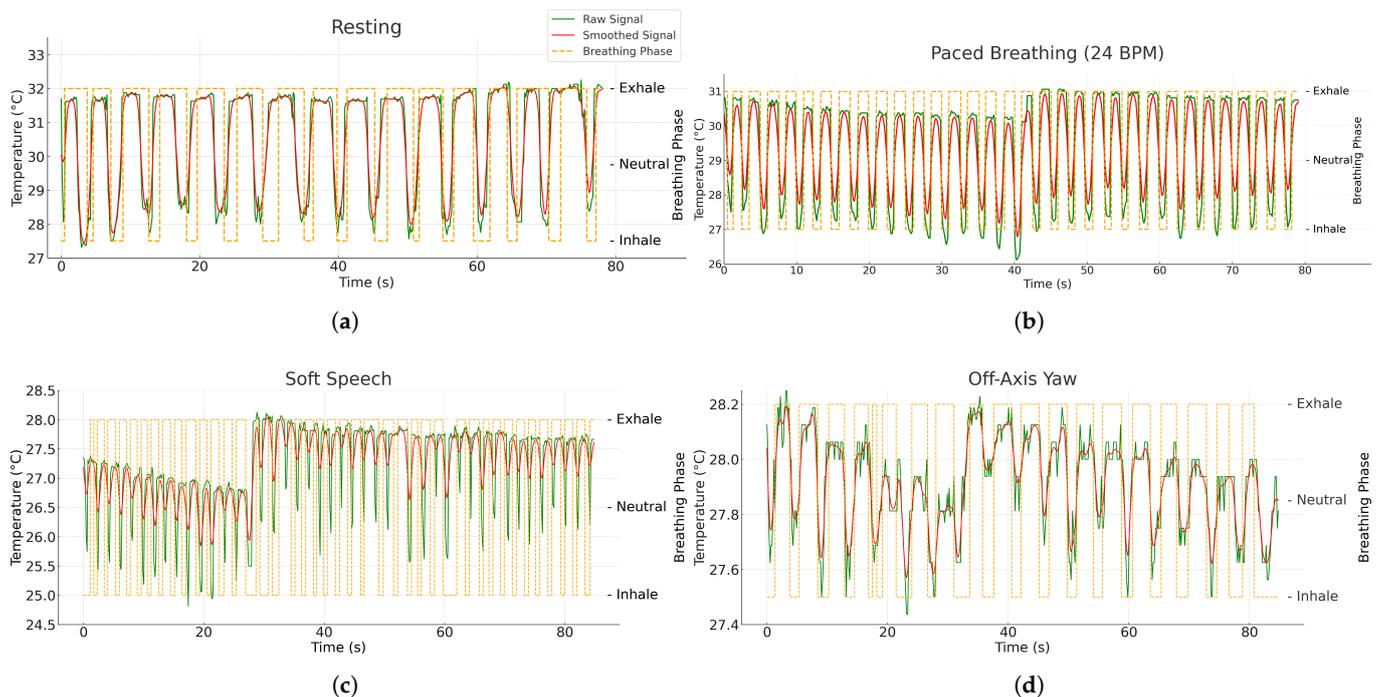


**Figure 14.** Representative thermal frames showing automatically detected nostril regions (green bounding boxes) across different participants and experimental conditions, including relaxed, off-axis, distance (2 m) and supine.

Once the nostril ROI is successfully detected, the system extracts the corresponding temperature signal to determine the breathing phases. Figure 15 presents representative nostril-temperature waveforms under four typical experimental conditions: resting, paced breathing (24 BPM), soft speech, and off-axis yaw. The green line indicates the raw temperature sequence, the red line depicts the smoothed and band-pass-filtered sig-

nal, and the orange-shaded regions denote exhalation phases identified by the adaptive MAD–hysteresis algorithm. During resting, as shown in Figure 15a, the thermal oscillations are smooth and periodic, reflecting stable nasal airflow. Under paced breathing in Figure 15b, the oscillation frequency increases in line with the metronome rhythm, confirming temporal consistency with the ground-truth reference. In soft-speech (Figure 15c) and off-axis (Figure 15d) scenarios, irregularities appear due to motion and partial ROI displacement; however, the system still successfully tracks the phase transitions, demonstrating robustness to moderate motion and physiological variability.

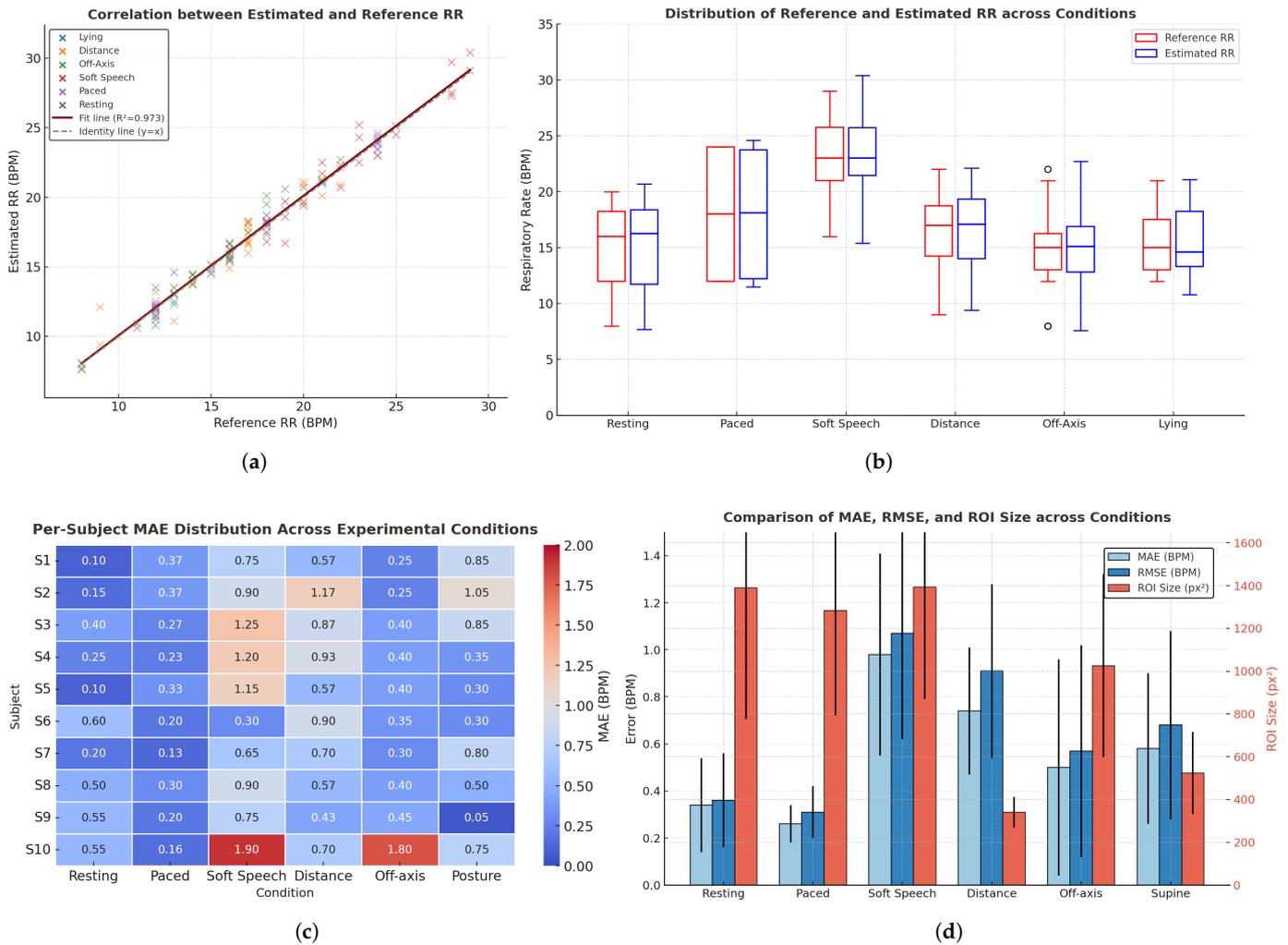
The respiratory-rate estimation experiment was conducted with ten healthy participants (age:  $33.3 \pm 4.38$  years). Each subject completed six breathing conditions described in Table 2, performed in real-time under different room layouts and lighting environments.



**Figure 15.** Representative breathing pattern signals extracted from the nostril ROI across four experimental conditions: (a) resting, (b) paced breathing, (c) soft speech, (d) off-axis yaw.

A summary of the results is presented in Table 3, which provides a per-subject breakdown of MAE, RMSE, and ROI pixel area across the six conditions, highlighting both inter-subject and condition-specific variability. Most participants exhibit stable performance during resting and paced breathing, whereas higher errors emerge during soft-speech and distance-related conditions due to motion artifacts and reduced signal-to-noise ratio (SNR). Notably, although the ROI area during soft speech remains larger than that in the distance condition, the estimation error is still considerably higher, suggesting that dynamic facial motion, rather than ROI size, is the primary factor contributing to performance degradation.

The correlation between the estimated and reference respiratory rates is shown in Figure 16a, demonstrating a strong linear relationship with  $R^2 = 0.973$ . These results indicate that the system can reliably estimate respiratory rate with minimal error across different conditions and sessions. To further examine the condition-wise performance, Figure 16b compares the distribution of estimated and reference respiratory rates using box plots, revealing close alignment across most conditions, with only slight deviations observed during soft-speech and off-axis orientations.



**Figure 16.** Summary of respiratory rate estimation results across all experimental conditions. (a) Correlation between estimated and reference respiratory rates, showing strong agreement across all breathing conditions. (b) Box-plot comparison between reference and estimated respiratory rates under different conditions, illustrating close alignment and low dispersion. (c) Per-subject MAE distribution across six experimental conditions, highlighting individual and condition-specific variability. (d) Overall MAE, RMSE, and ROI-size analysis per condition, showing accuracy degradation as camera distance increases due to reduced nostril-region pixels and weakened thermal contrast; notably, during soft speech, increased facial motion leads to higher estimation error despite a larger ROI area compared with the distance condition.

Moreover, Figure 16c illustrates the per-subject MAE distribution across the six conditions, highlighting both individual variability and condition-dependent performance. Estimation remains stable during resting and paced breathing, whereas larger errors appear during soft speech, off-axis yaw, and increased camera distance due to motion-induced disturbances and weakened thermal signal fidelity. Figure 16d summarizes the average MAE, RMSE, and ROI size per condition, showing that accuracy declines beyond 1.5 m as the nostril region becomes smaller and the thermal contrast diminishes. Interestingly, despite having a larger ROI area than the distance condition, the soft-speech condition exhibits higher estimation errors, reinforcing that facial dynamics, rather than ROI scale, are the dominant factor affecting accuracy.

**Table 3.** MAE, RMSE, and ROI pixel area (px<sup>2</sup>) of respiratory rate estimation per subject under six experimental conditions (pilot evaluation, N = 10).

Subj.	Resting			Paced			Soft Speech			Distance			Off-Axis			Posture		
	MAE	RMSE	px <sup>2</sup>	MAE	RMSE	px <sup>2</sup>	MAE	RMSE	px <sup>2</sup>	MAE	RMSE	px <sup>2</sup>	MAE	RMSE	px <sup>2</sup>	MAE	RMSE	px <sup>2</sup>
S1	0.1	0.1	2254	0.37	0.4	1847	0.75	0.99	1779	0.57	0.74	364	0.25	0.35	1216	0.85	1.33	808
S2	0.15	0.21	1742	0.37	0.46	2045	0.9	0.98	1888	1.17	1.81	493	0.25	0.29	1420	1.05	1.18	769
S3	0.4	0.41	809	0.27	0.32	785	1.25	1.57	709	0.87	1.14	286	0.4	0.5	831.5	0.85	0.91	544
S4	0.25	0.25	1390	0.23	0.35	1065	1.2	1.3	1528	0.93	1.02	367	0.4	0.5	831	0.35	0.38	500
S5	0.1	0.1	1401	0.33	0.42	1359	1.15	1.16	1984	0.57	0.70	273	0.4	0.41	1164	0.3	0.36	341
S6	0.6	0.6	2359	0.2	0.22	1738	0.3	0.36	1705	0.90	0.95	329	0.35	0.49	1897	0.3	0.31	597
S7	0.2	0.28	1102	0.13	0.14	846	0.65	0.65	861	0.70	0.79	261	0.3	0.42	777	0.8	1.06	301
S8	0.5	0.59	713	0.3	0.34	624	0.9	0.95	731	0.57	0.68	360	0.4	0.41	542	0.5	0.54	427
S9	0.55	0.57	1081	0.2	0.29	975	0.75	0.79	1046	0.43	0.55	317	0.45	0.51	935	0.05	0.07	496
S10	0.55	0.57	1055	0.16	0.17	1559	1.9	1.94	1696	0.70	0.72	352	1.8	1.82	644	0.75	0.87	446

Table 4 summarizes the average respiratory-rate estimation performance across all tested conditions, expressed as mean  $\pm$  SD for both MAE and RMSE. The results demonstrate that the system achieves consistently low errors across all scenarios, with the lowest MAE and RMSE observed during resting and paced breathing. Under conditions involving speech or posture change, the estimation error slightly increases, reflecting temperature fluctuations and ROI variation. For comparison, the included peak-based and FFT-based baseline methods show substantially higher errors across all conditions, confirming the advantage of the proposed adaptive approach. The overall error remains below 1 BPM on average, confirming clinically acceptable [52] performance for a lightweight thermal-based system operating on an embedded device.

**Table 4.** Respiratory rate estimation errors across all experimental conditions (mean  $\pm$  SD, N = 10). Baseline peak/FFT methods use one representative trial per subject due to their high sensitivity to artifacts, while the proposed method uses all recordings.

Condition	MAE (BPM)			RMSE (BPM)			ROI (px <sup>2</sup> )
	Proposed	Peak	FFT	Proposed	Peak	FFT	
Resting (Spontaneous)	0.34 $\pm$ 0.20	10.07 $\pm$ 5.48	1.51 $\pm$ 1.30	0.36 $\pm$ 0.20	11.24 $\pm$ 5.48	2.10 $\pm$ 1.30	1390 $\pm$ 614
Paced Breathing (Metronome)	0.26 $\pm$ 0.08	0.93 $\pm$ 0.65	5.11 $\pm$ 5.90	0.31 $\pm$ 0.11	1.12 $\pm$ 0.65	7.58 $\pm$ 5.90	1284 $\pm$ 491
Robustness (Soft Speech)	0.98 $\pm$ 0.43	4.39 $\pm$ 2.99	3.70 $\pm$ 3.92	1.07 $\pm$ 0.45	5.22 $\pm$ 2.99	5.24 $\pm$ 3.92	1393 $\pm$ 522
Distance (1.0–2.0 m)	0.74 $\pm$ 0.27	8.33 $\pm$ 5.19	5.76 $\pm$ 5.77	0.91 $\pm$ 0.37	9.68 $\pm$ 5.19	7.94 $\pm$ 5.78	340 $\pm$ 206
Off-axis Yaw ( $\pm$ 30°)	0.50 $\pm$ 0.46	10.72 $\pm$ 3.80	2.31 $\pm$ 3.79	0.57 $\pm$ 0.45	11.30 $\pm$ 3.80	3.78 $\pm$ 3.79	1026 $\pm$ 428
Posture (Supine)	0.58 $\pm$ 0.32	6.31 $\pm$ 4.04	3.47 $\pm$ 3.94	0.68 $\pm$ 0.40	7.38 $\pm$ 4.04	5.10 $\pm$ 3.94	523 $\pm$ 192
Overall	0.57 $\pm$ 0.36	6.79 $\pm$ 3.86	3.64 $\pm$ 3.88	0.64 $\pm$ 0.42	7.58 $\pm$ 3.86	5.48 $\pm$ 3.88	–

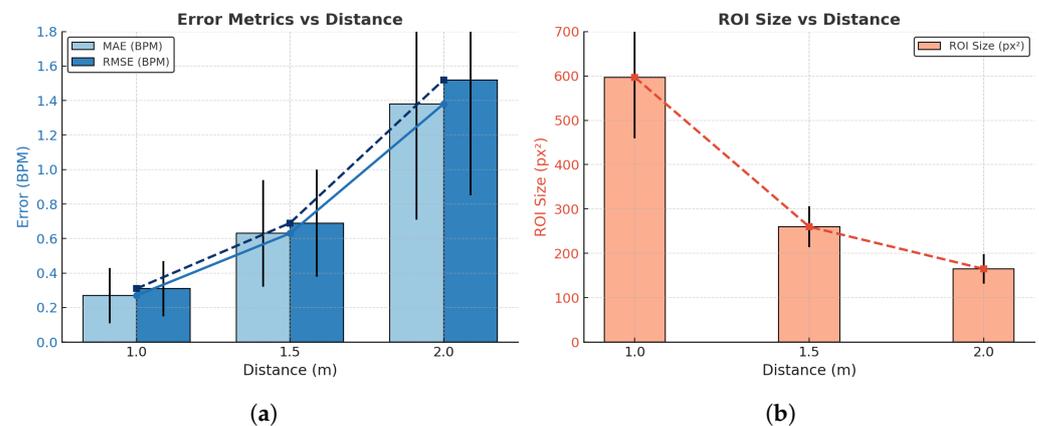
To further investigate the influence of distance on detection scale and estimation accuracy, Table 5 reports the mean  $\pm$  SD of MAE, RMSE, and ROI size across the three measurement distances. The results show a substantial reduction in detected ROI area as the camera moves farther away—from 597 px<sup>2</sup> at 1.0 m to just 165 px<sup>2</sup> at 2.0 m—representing a 72% decrease in spatial sampling. This loss of pixel coverage directly diminishes thermal contrast and reduces signal amplitude, leading to higher estimation errors at extended distances. The increasing error trend therefore, aligns with the shrinking ROI size, confirming that reduced spatial resolution limits the system’s ability to capture subtle nostril temperature variations. Nevertheless, within the practical monitoring range of 1.0–1.5 m, estimation performance remains stable, with MAE values below 0.7 BPM.

**Table 5.** Effect of distance on respiratory rate estimation accuracy and ROI size (mean  $\pm$  SD,  $N = 10$ ).

Distance (m)	MAE (BPM)	RMSE (BPM)	ROI Size (px <sup>2</sup> )	Observation *
1.0	0.27 $\pm$ 0.16	0.31 $\pm$ 0.16	597 $\pm$ 138	Clear nostril region, distinct thermal contrast
1.5	0.63 $\pm$ 0.31	0.69 $\pm$ 0.31	260 $\pm$ 46	Reduced contrast, smaller ROI
2.0	1.38 $\pm$ 0.67	1.52 $\pm$ 0.67	165 $\pm$ 33	Weak contrast, partial pixel loss

\* Observations are based on visual inspection of thermal contrast and ROI clarity during recordings.

A more detailed visualization of the distance–accuracy relationship is provided in Figure 17. As shown in Figure 17a, the detection error increases markedly, from 0.27 BPM MAE and 0.31 BPM RMSE at 1 m to 1.38 BPM MAE and 1.52 BPM RMSE at 2 m, while the corresponding ROI area decreases from approximately 597 px<sup>2</sup> to 165 px<sup>2</sup> (Figure 17b). This consistent trend reinforces that distance-induced loss of spatial detail is the primary factor driving performance degradation.



**Figure 17.** The effect of measurement distance on respiratory rate estimation performance. (a) shows the variation of error metrics (MAE and RMSE) with increasing distance, where solid bars denote MAE and dashed lines denote RMSE; (b) presents the corresponding change in ROI size across different measurement distances, where the dashed line highlights the overall trend.

To contextualize these device-level gains, the proposed system is compared with recent contactless respiratory-rate estimation studies. Table 6 summarizes relevant methods using thermal imaging, highlighting differences in hardware, ROI selection, tracking strategy, estimation algorithm, accuracy, runtime feasibility, and overall contribution. The developed system achieves markedly lower estimation errors than most recent thermal-based approaches, attaining a mean absolute error of  $0.57 \pm 0.36$  BPM and an RMSE of  $0.64 \pm 0.42$  BPM across diverse conditions, including speech, head rotation, and distances up to 2.0 m. These results fall well within the commonly accepted clinical tolerance for respiratory-rate monitoring (error  $< 2$  BPM) [52], corresponding to an average deviation below 1 BPM. The system thus delivers clinically relevant accuracy using a lightweight  $256 \times 192$  thermal camera.

Compared to deep learning–based approaches, which achieve good accuracy but require high-resolution cameras, complex models, and non-real-time post-processing [19,26], the proposed system emphasizes lightweight edge deployment with minimal computational cost. Cross-modality solutions that fuse RGB and thermal data have demonstrated clinical viability but introduce additional sensor complexity and are not optimized for embedded platforms [22]. Meanwhile, traditional spectral methods achieve competitive accuracy but typically lack automated ROI localization and real-time performance [52–54]. In

contrast, the proposed system is the first to integrate nostril-specific ROI tracking, adaptive MAD-hysteresis phase detection, and IBI validation within a real-time, edge-deployable framework. This combination enables robustness to motion and viewpoint variation while achieving state-of-the-art accuracy and real-time operation suitable for continuous home monitoring.

**Table 6.** Comparative summary of recent contactless respiratory rate (RR) estimation methods using thermal imaging. Each method is evaluated based on subjects/conditions, camera specifications, setup (including ROI and method), accuracy, runtime feasibility, and novelty. The proposed system emphasizes nostril-focused ROI, IBI validation, and real-time edge deployment.

Study	Subjects & Conditions	Camera Spec & Setup (ROI + Method)	Accuracy (BPM)	Real-Time	Contribution
Ours	10 adults; six 60-s sets (resting, paced 12/18/24 BPM, soft speech, distance 1–2 m, yaw $\pm 30^\circ$ , posture supine)	TOPDON TC001 (256 $\times$ 192); RO: Ithermal YOLOv8n (even frames, $s = 10$ ) + Kalman tracking on skipped frames; band-pass 0.08–0.7 Hz; adaptive MAD-hysteresis phase detection + IBI validation	MAE (mean $0.57 \pm 0.36$ ); RMSE (mean $0.64 \pm 0.42$ )	Yes	Thermal-based YOLO detector with Kalman tracking; adaptive MAD-hysteresis phase and IBI validation
Gioia et al. [19]	30 adults; 5-min tasks (rest, Stroop, emotion); RR 9–30	FLIR T640 (640 $\times$ 480); ROI: upper lip/nose (manual); 3D-CNN end-to-end regression	$R^2 \approx 0.61$ (no MAE/RMSE)	No	Feasibility of end-to-end deep learning directly from thermal video
Mozafari et al. [26]	22 adults; sitting/standing $\times$ mask/no-mask, 90 s	FLIR T650sc (640 $\times$ 480); ROI: full face (DeTr); 3D-CNN + BiLSTM with correlation loss	MAE $1.6 \pm 0.4$	Yes	Deep learning robust to mask & posture; real-time feasibility focus
Maurya et al. [22]	14 adults (rest, talking, variable); 8 neonates (NICU)	FLIR-E60 (320 $\times$ 240) + Logitech C922 RGB (960 $\times$ 720); ROI: nose–mouth (from RGB mapped to thermal); Hampel+MA+BP filtering; CZT spectral analysis	Adults: MAE 0.10–1.8; Neonates: MAE $\approx 1.5$	No	Cross-modality ROI mapping; validated adults & neonates
Takahashi et al. [52]	7 adults; paced 15–30 BPM	FLIR Boson 320 (320 $\times$ 256); ROI: face subregions (scored by RQI); YOLOv3 + RQI; FFT on best region	MAE 0.66; LoA $\pm 2$ BPM	No	ROI quality index (RQI) for automated subregion selection
Pereira et al. [53]	12 adults (rest, pathological); 8 neonates (NICU)	InfraTec VarioCAM HD (1024 $\times$ 768); ROI: full face (multi-grid, black-box); adaptive spectral analysis (autocorr, AMDF, peak detection)	RMSE 0.31 (rest), 3.27 (varied), 4.15 (neonates)	No	First NICU validation; black-box ROI without anatomical landmark
Nakai et al. [54]	11 healthy adults; seated at 1 m distance (lab environment)	FLIR A315 (320 $\times$ 240, 60 fps); manually defined nose & shoulder ROIs; dual-signal extraction (thermal variation + shoulder motion); band-pass filtering + autocorrelation/FFT for RR estimation	$r = 0.83$ vs. belt; MAE $\approx 2.4$ ; RMSE $\approx 2.4$	No	Dual-ROI thermal approach combining nasal temperature and shoulder motion

## 5. Discussion

The experimental results demonstrate that the proposed thermal-based system, which integrates a YOLO-based nostril detector executed on every second frame with Kalman prediction and an adaptive breathing-phase and IBI validation module, enables accurate and robust respiratory-rate estimation entirely on an embedded edge device. The frame-skipping strategy effectively reduces computational demand without compromising ROI continuity, while the time-domain phase logic, based on median and MAD thresholds with hysteresis and short-segment consolidation, maintains stable breathing-phase labeling under thermal drift and motion-induced noise. To strengthen the evaluation, two baseline methods, such as peak detection and FFT-based spectral analysis, were implemented and tested on the same real-time data collected in the study. As shown in Table 4, these baselines exhibit substantially higher MAE and RMSE across all conditions, confirming the

advantage of the proposed method. Although several recently published thermal-based respiratory-rate estimation methods exist, direct comparison was not feasible. Most rely on substantially higher-resolution thermal sensors, computationally intensive 3D-CNN or transformer architectures, or RGB–thermal fusion pipelines that cannot be reproduced on the low-resolution dataset used in this study or deployed on embedded hardware.

In terms of privacy, the thermal modality used here does not capture facial texture, identity cues, or personally identifiable imagery. The  $256 \times 192$  thermal frames contain only coarse temperature gradients, and the respiratory pipeline operates exclusively on a small nostril-level ROI, further reducing the possibility of re-identification. Although a formal privacy-impact assessment was not mandated for this pilot study, the sensing modality is inherently privacy-preserving compared with RGB-based approaches. When contrasted with genuinely anonymous alternatives such as radar or acoustic sensing, thermal imaging provides a favorable balance between privacy and spatial specificity: radar and acoustic systems offer strong anonymity but often exhibit reduced spatial precision, susceptibility to multipath or ambient noise, and difficulty maintaining stable anatomical anchoring for breath extraction [55,56]. By leveraging non-textured thermal data while retaining reliable nostril localization, the proposed approach achieves privacy-aware respiratory monitoring without compromising estimation accuracy.

Across all evaluated conditions, respiratory-rate estimation remained consistently accurate, with MAE values ranging from 0.34 to 0.98 BPM and RMSE values between 0.36 to 1.07 BPM across ten participants with an overall average of  $0.57 \pm 0.36$  BPM with RMSE  $0.64 \pm 0.42$  BPM. Errors were lowest during resting and paced-breathing trials, and increased modestly under soft-speech and posture variations due to mouth motion and partial ROI displacement. Distance and off-axis tests showed moderate increases in error, indicating that the system maintains reliable estimation up to approximately 1.5 m, even with a low-resolution thermal camera sensor. These results confirm that accurate, real-time respiratory-rate monitoring can be achieved using an adaptive, privacy-preserving thermal camera operating entirely on an edge platform. The experimental findings also demonstrate the adaptive nature of the proposed system across diverse conditions. Rather than relying on fixed thresholds or static parameters, the MAD-based phase logic continuously adjusts to variations in signal amplitude and noise, while the hysteresis and consolidation mechanisms ensure stable breathing-phase transitions. These adaptive behaviors collectively enable consistent performance under different breathing patterns and motion scenarios without manual recalibration.

A closer examination of failure modes provides further insight into the system's behavior under challenging scenarios. During soft-speech condition, the primary source of degradation arose not only from general facial motion but also from rapid upper-lip deformation and transient nostril occlusion caused by articulation. These movements introduce abrupt thermal discontinuities within the ROI, reducing local temporal coherence and lowering the effective signal-to-noise ratio (SNR) of the extracted temperature waveform. A similar failure trend is observed with increasing camera distance, where the nostril region shrinks from  $597 \text{ px}^2$  at 1.0 m to  $165 \text{ px}^2$  at 2.0 m. This reduction leads to diminished thermal gradient resolution, smaller oscillation amplitudes, and greater sensitivity to pixel-level quantization noise. Together, these analyses clarify the mechanisms underlying soft-speech- and distance-related performance degradation, complementing the quantitative results in Table 4.

Compared with existing methods summarized in Table 6, the primary advantage of the proposed system lies in its fully automated, real-time processing at minimal computational cost. Solely deep learning-based approaches can achieve strong performance but typically depend on high-resolution cameras and computationally intensive models, often requiring

offline post-processing [19,26]. In contrast, conventional spectral methods can run on simpler hardware yet usually lack automated ROI localization and are sensitive to motion and baseline temperature drift [52,53]. The present design achieves both efficiency and stability by performing YOLO detections on thermal imagery, transferring the ROI via calibrated alignment, applying Kalman prediction on skipped frames, and estimating the breathing pattern directly in the time domain. This architecture maintains robustness to motion and noise while remaining fully compatible with real-time embedded execution.

A direct comparison with recent studies further contextualizes system performance. Mozafari et al. [26] reported an MAE of approximately 1.6 BPM using a  $640 \times 480$  thermal camera with a 3D-CNN + BiLSTM model, while Nakai et al. [54] reported MAE values around 2.4 BPM using dual thermal ROIs. Gioia et al. [19] achieved an  $R^2$  of roughly 0.10 with high-resolution imagery and offline 3D-CNN regression. Classical FFT/CZT-based approaches typically achieve MAE between 0.66 and 1.8 BPM under controlled conditions but depend on manual or semi-automated ROI selection [22]. In contrast, the proposed system delivers competitive accuracy using a lower-resolution  $256 \times 192$  sensor while maintaining fully automated ROI tracking and real-time embedded execution. Furthermore, most prior thermal-based studies evaluated only one or two controlled breathing conditions, whereas the proposed system was validated across six diverse scenarios, including speech, off-axis rotation, posture variation, and distances up to 2.0 m, demonstrating robustness under a wider range of real-world variations.

Regarding model validity, the respiratory-rate estimation module does not involve any data-driven training and therefore cannot overfit to the subjects. All processing parameters, including band-pass filter settings, MAD-based thresholds, and hysteresis rules, were fixed in advance and applied identically to all participants. The YOLO nostril detector, the only trained component in the pipeline, was trained on an independent thermal dataset (7958 annotated frames) that did not include any of the subjects used in the *RR* evaluation. Ground-truth *RR* was obtained through dual-rater manual counting, and system accuracy was assessed using MAE and RMSE across all six experimental conditions.

Despite the advantages of the proposed method, several limitations were identified. The system requires a clearly visible nostril region to estimate the respiratory rate accurately. When the nostrils are covered, exhibit low thermal contrast, or move outside the camera's field of view, such as when subjects wear masks or turn their heads excessively, the system may fail to produce valid respiratory rate estimates because no stable thermal signal can be extracted. In addition, accuracy decreases as the camera-to-subject distance increases. Beyond approximately 1.5–2.0 m, the nostril region becomes very small in the thermal frame, the average ROI area decreases from about  $597 \text{ px}^2$  to  $165 \text{ px}^2$ , reducing temperature contrast and making the signal more susceptible to noise and minor tracking errors. Moreover, the proposed method is effective only when breathing occurs predominantly through the nasal pathway. During diaphragmatic or abdominal breathing, where nasal airflow is minimal, the thermal contrast around the nostrils becomes negligible, leading to weak or undetectable respiratory oscillations. Furthermore, when the ambient temperature is high, the subject's facial temperature increases, diminishing the thermal contrast and making face or nostril detection unreliable.

As this work represents a feasibility study to demonstrate whether a low-resolution thermal camera can reliably provide non-invasive respiratory-rate monitoring in real time, the evaluation was intentionally limited to ten healthy adults within a narrow age range (mean age  $33.3 \pm 4.38$  years). Consequently, elderly adults or individuals with respiratory conditions such as COPD, asthma, or sleep apnoea were not included, resulting in limited clinical validation and reduced population diversity. Future work will conduct broader

clinical validation involving these patient groups, as well as individuals at risk of long-lie incidents, to ensure that the system performs reliably across diverse real-world populations.

To support these clinical applications, the system also requires further technical enhancements to ensure reliable respiration monitoring under more challenging real-world conditions. Future improvements may include integrating a low-power radar module to complement the thermal sensing, enabling respiration-phase estimation even when nasal airflow is weak or partially occluded. Beyond nasal-based measurements, future work will also explore extracting respiratory information from micro-motions of the shoulder or abdomen and identifying mouth-breathing episodes. Furthermore, fusing the proposed respiratory-rate estimator with the previously developed long-lie detection system [57] may improve reliability and reduce false alarms across varying distances, postures, and environmental conditions. This multi-modal integration would move the system toward more robust and clinically relevant continuous home monitoring.

## 6. Conclusions

This paper introduces an adaptive, fully automated, and privacy-preserving respiratory-rate monitoring system based on thermal imaging, designed for real-time execution on embedded edge hardware. The framework integrates a lightweight thermal-specific YOLO-based nostril detector, a detector-centric frame-skipping strategy with Kalman prediction for stable ROI continuity, and an adaptive median-MAD hysteresis algorithm with consolidation and IBI validation for robust time-domain respiration analysis.

Across six experimental conditions, including speech, off-axis rotation, posture variation, and distances up to 2.0 m, the system achieved an average MAE of  $0.57 \pm 0.36$  BPM and RMSE of  $0.64 \pm 0.42$  BPM, demonstrating that accurate and reliable respiratory-rate estimation is achievable using a compact  $256 \times 192$  thermal sensor operating fully on a low-power embedded platform. The adaptive signal-processing pipeline consistently adjusted to variations in breathing amplitude, rhythm, and motion-induced disturbances without requiring manual recalibration. Notably, the achieved accuracy falls well within clinically acceptable tolerances for respiratory-rate monitoring, reinforcing its suitability for practical deployment in home or long-term monitoring environments.

Future work will include expanding the participant population and range of respiratory scenarios, improving resilience against occlusion and abdomen- or mouth-dominant breathing through the integration of a low-power radar module, and embedding the proposed respiratory module into an existing long-lie detection system. This multi-modal fusion of thermal and physiological information aims to enhance robustness, reduce false alarms, and support continuous, privacy-preserving home monitoring.

**Author Contributions:** R.A.: Conceptualization, Data Curation, Formal Analysis, Methodology, Software, Validation, Visualization, Writing—original draft. A.F.: Methodology, Supervision, Validation, Writing—review and editing. S.-Q.X.: Supervision, Methodology, Writing—review and editing. Z.Z.: Conceptualization, Supervision, Validation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia [Grant No. LOG-13094/LPDP.3/2024].

**Institutional Review Board Statement:** All procedures involving human participants were approved by the Faculty Research Ethics Committee (FREC) for Engineering and Physical Sciences (EPS), University of Leeds (Application No. 1499). Informed consent was obtained from all participants.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

**Acknowledgments:** The authors would like to thank all participants who were recruited for all the experiments of respiratory rate estimation.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

YOLO	You Only Look Once
ROI	Region of Interest
IIR	Infinite Impulse Response
IBI	Inter-Breath Interval
RR	Respiratory Rate
3D-CNNs	Three-Dimensional Convolutional Neural Networks
SSD	Single Shot MultiBox Detector
YUV	Luminance (Y) and chrominance components (U: blue-difference, V: red-difference)
C2f	Cross Stage Partial with 2 convolutions and fusion
BPM	Breaths Per Minute
GUI	Graphical User Interface
GPU	Graphics Processing Unit
CPU	Central Processing Unit
DFL	Distribution Focal Loss
PR	Precision–Recall
BiLSTM	Bidirectional Long Short-Term Memory
CZT	Chirp Z-Transform
RQI	Region Quality Index
FFT	Fast Fourier Transform
AMDF	Average Magnitude Difference Function
NICU	Neonatal Intensive Care Unit
FPS	Frames Per Second
RGB	Red, Green, Blue color space
MAD	Median Absolute Deviation
EMA	Exponential Moving Average
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
SNR	signal-to-noise ratio
p05	5th percentile
p95	95th percentile
MSB	most significant byte
LSB	least significant byte
Std Dev	Standard deviation
COPD	Chronic Obstructive Pulmonary Disease

## References

1. Drummond, G.; Fischer, D.; Arvind, D. Current clinical methods of measurement of respiratory rate give imprecise values. *ERJ Open Res.* **2020**, *6*, 00023–2020. [[CrossRef](#)]
2. Tobin, M.J. Breathing pattern analysis. *Intensive Care Med.* **2005**, *18*, 193–201. [[CrossRef](#)] [[PubMed](#)]
3. Ashe, W.B.; McNamara, B.D.; Patel, S.M.; Shanno, J.N.; Innis, S.E.; Hochheimer, C.J.; Barros, A.J.; Williams, R.D.; Ratcliffe, S.J.; Moorman, J.; et al. Kinematic signature of high risk labored breathing revealed by novel signal analysis. *Sci. Rep.* **2024**, *14*, 27794. [[CrossRef](#)]
4. Rivas, E.; López-Baamonde, M.; Sanahuja, J.; del Rio, E.; Ramis, T.; Recasens, A.; López, A.; Arias, M.; Kampakis, S.; Lauteslager, T.; et al. Early detection of deterioration in COVID-19 patients by continuous ward respiratory rate monitoring: A pilot prospective cohort study. *Front. Med.* **2023**, *10*, 1243050. [[CrossRef](#)]
5. Peters, G.; Peelen, R.; Gilissen, V.; Koning, M.; Harten, W.; Doggen, C. Detecting patient deterioration early using continuous heart rate and respiratory rate measurements in hospitalized COVID-19 patients. *J. Med. Syst.* **2023**, *47*, 12. [[CrossRef](#)]

6. Yadav, A.; Dandu, H.; Parchani, G.; Chokalingam, K.; Kadambi, P.; Mishra, R.; Jahan, A.; Teboul, J.; Latour, J. Early detection of deteriorating patients in general wards through continuous contactless vital signs monitoring. *Front. Med. Technol.* **2024**, *6*, 1436034. [[CrossRef](#)]
7. McCartan, T.; Worrall, A.; Conluain, R.; Alaya, F.; Mulvey, C.; MacHale, E.; Brennan, V.; Lombard, L.; Walsh, J.; Murray, M.; et al. The effectiveness of continuous respiratory rate monitoring in predicting hypoxic and pyrexia events: A retrospective cohort study. *Physiol. Meas.* **2021**, *42*, 065005. [[CrossRef](#)] [[PubMed](#)]
8. Ryyänen, O.P.; Kivelä, S.; Honkanen, R.; Laippala, P. Falls and lying helpless in the elderly. *Z. Gerontol.* **1992**, *25*, 278–282. [[PubMed](#)]
9. Fleming, J.; Brayne, C. Inability to get up after falling, subsequent time on floor, and summoning help: Prospective cohort study in people over 90. *BMJ* **2008**, *337*, a2227. [[CrossRef](#)]
10. Kubitzka, J.; Schneider, I.T.; Reuschenbach, B. Concept of the term long lie: A scoping review. *Eur. Rev. Aging Phys. Act.* **2023**, *20*, 16. [[CrossRef](#)]
11. Massaroni, C.; Nicolò, A.; Presti, D.; Sacchetti, M.; Silvestri, S.; Schena, E. Contact-Based Methods for Measuring Respiratory Rate. *Sensors* **2019**, *19*, 908. [[CrossRef](#)] [[PubMed](#)]
12. Costanzo, I.; Sen, D.; Rhein, L.; Guler, U. Respiratory Monitoring: Current State of the Art and Future Roads. *IEEE Rev. Biomed. Eng.* **2020**, *15*, 103–121. [[CrossRef](#)]
13. Liu, H.; Allen, J.; Zheng, D.; Chen, F. Recent development of respiratory rate measurement technologies. *Physiol. Meas.* **2019**, *40*, 07TR01. [[CrossRef](#)] [[PubMed](#)]
14. Vitazkova, D.; Foltan, E.; Kosnacova, H.; Micjan, M.; Donoval, M.; Kuzma, A.; Kopani, M.; Vavrinsky, E. Advances in Respiratory Monitoring: A Comprehensive Review of Wearable and Remote Technologies. *Biosensors* **2024**, *14*, 90. [[CrossRef](#)]
15. Xia, Z.; Shi, M.M.H.; Li, Y.; Inan, O.T.; Zhang, Y. The Delineation of Fiducial Points for Non-Contact Radar Seismocardiogram Signals Without Concurrent ECG. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1031–1040. [[CrossRef](#)] [[PubMed](#)]
16. Jalaja, A.A.; Kavitha, M. Contactless face video based vital signs detection framework for continuous health monitoring using feature optimization and hybrid neural network. *Biotechnol. Bioeng.* **2024**, *121*, 1190–1214. [[CrossRef](#)]
17. Chian, D.-M.; Wen, C.-K.; Wang, C.-J.; Hsu, M.-H.; Wang, F.-K. Vital Signs Identification System With Doppler Radars and Thermal Camera. *IEEE Trans. Biomed. Circuits Syst.* **2022**, *16*, 153–167. [[CrossRef](#)]
18. Negishi, T.; Abe, S.; Matsui, T.; Liu, H.; Kurosawa, M.; Kirimoto, T.; Sun, G. Contactless Vital Signs Measurement System Using RGB-Thermal Image Sensors and Its Clinical Screening Test on Patients with Seasonal Influenza. *Sensors* **2020**, *20*, 2171. [[CrossRef](#)]
19. Gioia, F.; Pura, F.; Greco, A.; Piga, D.; Merla, A.; Forgione, M. Contactless Estimation of Respiratory Frequency Using 3D-CNN on Thermal Images. *IEEE J. Biomed. Health Inform.* **2025**, *29*, 7387–7395. [[CrossRef](#)]
20. Yin, C.; Wang, G.; Xie, Y.; Tu, J.; Sun, W.; Kong, X.; Guo, X.; Zhang, D. Separated Respiratory Phases for In Vivo Ultrasonic Thermal Strain Imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2022**, *69*, 1219–1229. [[CrossRef](#)]
21. Alves, R.; van Meulen, F.; Overeem, S.; Zinger, S.; Stuijk, S. Thermal Cameras for Continuous and Contactless Respiration Monitoring. *Sensors* **2024**, *24*, 8118. [[CrossRef](#)] [[PubMed](#)]
22. Maurya, L.; Zwiggelaar, R.; Chawla, D.; Mahapatra, P. Non-contact respiratory rate monitoring using thermal and visible imaging: A pilot study on neonates. *J. Clin. Monit. Comput.* **2022**, *37*, 815–828. [[CrossRef](#)]
23. Jakkaew, P.; Onoye, T. Non-Contact Respiration Monitoring and Body Movements Detection for Sleep Using Thermal Imaging. *Sensors* **2020**, *20*, 6307. [[CrossRef](#)]
24. Hurtado, D.; Abusleme, A.; Chávez, J.A.P. Non-invasive continuous respiratory monitoring using temperature-based sensors. *J. Clin. Monit. Comput.* **2019**, *34*, 223–231. [[CrossRef](#)]
25. Jisha, S.; Jayanthi, T. Non-Contact methods for assessment of respiratory parameters. In Proceedings of the 2023 International Conference on Innovations in Engineering and Technology (ICIET), Muvattupuzha, India, 13–14 July 2023; pp. 1–6. [[CrossRef](#)]
26. Mozafari, M.; Law, A.J.; Goubran, R.A.; Green, J.R. Respiratory Rate Estimation from Thermal Video Data Using Spatio-Temporal Deep Learning. *Sensors* **2024**, *24*, 6386. [[CrossRef](#)]
27. Choi, J.; Oh, K.; Kwon, O.; Kwon, J.; Kim, J.; Yoo, S. Non-Contact Respiration Rate Measurement From Thermal Images Using Multi-Resolution Window and Phase-Sensitive Processing. *IEEE Access* **2023**, *11*, 112706–112718. [[CrossRef](#)]
28. Xu, Y.; Khan, T.M.; Song, Y.; Meijering, E. Edge deep learning in computer vision and medical diagnostics: A comprehensive survey. *Artif. Intell. Rev.* **2025**, *58*, 93. [[CrossRef](#)]
29. Rancea, A.; Anghel, I.; Cioara, T. Edge computing in healthcare: Innovations, opportunities, and challenges. *Future Internet* **2024**, *16*, 329. [[CrossRef](#)]
30. Pankaj, A.; Kumar, A.; Kumar, M.; Komaragiri, R. Edge-Based Computation of Super-Resolution Superlet Spectrograms for Real-Time Estimation of Heart Rate Using an IoMT-Based Reference-Signal-Less PPG Sensor. *IEEE Internet Things J.* **2024**, *11*, 8647–8657. [[CrossRef](#)]
31. Kumar, R.H.; Rajaram, B. Design and Simulation of an Edge Compute Architecture for IoT-Based Clinical Decision Support System. *IEEE Access* **2024**, *12*, 45456–45474. [[CrossRef](#)]

32. Kanungo, P. Edge computing in healthcare: Real-time patient monitoring systems. *World J. Adv. Eng. Technol. Sci.* **2025**, *15*, 1–9. [[CrossRef](#)]
33. Catalina, L.; Doru, A.; Călin, C. The use of thermographic techniques and analysis of thermal images to monitor the respiratory rate of premature new-borns. *Case Stud. Therm. Eng.* **2021**, *25*, 100926. [[CrossRef](#)]
34. Alva, R.; Talasila, V.; Tv, S.; Umar, S.A.R. Estimation of Respiratory Rate Using Thermography. In Proceedings of the 2024 5th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 4–5 October 2024; pp. 154–159. [[CrossRef](#)]
35. Luca, C.; Andritoi, D.; Corciova, C. Monitoring the Respiratory Rate in the Premature Newborn by Analyzing Thermal Images. *Nov. Perspect. Eng. Res.* **2022**, *5*, 129–143. [[CrossRef](#)]
36. Shu, S.; Liang, H.-R.; Zhang, Y.; Zhang, Y.; Yang, Z. Non-contact measurement of human respiration using an infrared thermal camera and the deep learning method. *Meas. Sci. Technol.* **2022**, *33*, 075202. [[CrossRef](#)]
37. Lazri, Z.; Zhu, Q.; Chen, M.; Wu, M.; Wang, Q. Detecting Essential Landmarks Directly in Thermal Images for Remote Body Temperature and Respiratory Rate Measurement With a Two-Phase System. *IEEE Access* **2022**, *10*, 39080–39094. [[CrossRef](#)]
38. Hu, M.; Zhai, G.; Li, D.; Fan, Y.; Duan, H.; Zhu, W.; Yang, X. Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation. *PLoS ONE* **2018**, *13*, e0190466. [[CrossRef](#)]
39. Kowalczyk, N.; Rumiński, J. Respiratory Rate Estimation Based on Detected Mask Area in Thermal Images. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada, 17–24 June 2023; pp. 6042–6051. [[CrossRef](#)]
40. Kwon, J.; Kwon, O.; Oh, K.; Kim, J.; Yoo, S. Breathing-Associated Facial Region Segmentation for Thermal Camera-Based Indirect Breathing Monitoring. *IEEE J. Transl. Eng. Health Med.* **2023**, *11*, 505–514. [[CrossRef](#)]
41. Cho, Y.; Julier, S.; Marquardt, N.; Bianchi-Berthouze, N. Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging. *Biomed. Opt. Express* **2017**, *8*, 4480–4503. [[CrossRef](#)] [[PubMed](#)]
42. Chen, D.-Y.; Lai, J. HHT-based remote respiratory rate estimation in thermal images. In Proceedings of the 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Kanazawa, Japan, 26–28 June 2017; pp. 263–268. [[CrossRef](#)]
43. Kwaśniewska, A.; Szankin, M.; Rumiński, J.; Kaczmarek, M. Evaluating Accuracy of Respiratory Rate Estimation from Super Resolved Thermal Imagery. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2744–2747. [[CrossRef](#)]
44. Romano, C.; Innocenti, L.; Schena, E.; Sacchetti, M.; Nicolò, A.; Massaroni, C. A Signal Quality Index for Improving the Estimation of Breath-by-Breath Respiratory Rate During Sport and Exercise. *IEEE Sens. J.* **2023**, *23*, 31250–31258. [[CrossRef](#)]
45. Sabz, M.; MacLean, J.; Martin, A.R.; Rouhani, H. Characterization of Wearable Respiratory Sensors for Breathing Parameter Measurements. *IEEE Sens. J.* **2024**, *24*, 32283–32290. [[CrossRef](#)]
46. Cheng, J.; Liu, R.; Li, J.; Song, R.; Liu, Y.; Chen, X. Motion-Robust Respiratory Rate Estimation From Camera Videos via Fusing Pixel Movement and Pixel Intensity Information. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 4008611. [[CrossRef](#)]
47. Wang, W.; den Brinker, A.C. Algorithmic insights of camera-based respiratory motion extraction. *Physiol. Meas.* **2022**, *43*, 075004. [[CrossRef](#)] [[PubMed](#)]
48. Romano, C.; Schena, E.; Silvestri, S.; Massaroni, C. Non-Contact Respiratory Monitoring Using an RGB Camera for Real-World Applications. *Sensors* **2021**, *21*, 5126. [[CrossRef](#)]
49. Yin, J.; Wu, F.; Su, H.; Huang, P.; Qixuan, Y. Improvement of SAM2 algorithm based on Kalman filtering for long-term video object segmentation. *Sensors* **2025**, *25*, 4199. [[CrossRef](#)]
50. Fernandez, I.C.; Magpantay, P.; Rosales, M.D.; Hizon, J.R.E. Review of Kalman filters in multiple object tracking algorithms. In Proceedings of the 2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS), London, UK, 29–31 July 2024; pp. 1–4. [[CrossRef](#)]
51. Kallioinen, N.; Hill, A.; Christofidis, M.; Horswill, M.; Watson, M. Quantitative systematic review: Sources of inaccuracy in manually measured adult respiratory rate data. *J. Adv. Nurs.* **2020**, *77*, 98–124. [[CrossRef](#)]
52. Takahashi, Y.; Gu, Y.; Nakada, T.; Abe, R.; Nakaguchi, T. Estimation of Respiratory Rate from Thermography Using Respiratory Likelihood Index. *Sensors* **2021**, *21*, 4406. [[CrossRef](#)]
53. Pereira, C.B.; Yu, X.; Goos, T.; Reiss, I.; Orlikowsky, T.; Heimann, K.; Venema, B.; Blazek, V.; Leonhardt, S.; Teichmann, D. Noncontact Monitoring of Respiratory Rate in Newborn Infants Using Thermal Imaging. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 1105–1114. [[CrossRef](#)]
54. Nakai, K.; Kurosawa, M.; Kirimoto, T.; Matsui, T.; Abe, S.; Suzuki, S.; Sun, G. Performance enhancement of thermal image analysis for noncontact cardiopulmonary signal extraction. *Infrared Phys. Technol.* **2024**, *138*, 105244. [[CrossRef](#)]
55. Yang, Z.; Liu, Y.; Yang, H.; Shi, J.; Hu, A.; Xu, J.; Zhuge, X.; Miao, J. Noncontact Breathing Pattern Monitoring Using a 120 GHz Dual Radar System with Motion Interference Suppression. *Biosensors* **2025**, *15*, 486. [[CrossRef](#)] [[PubMed](#)]

56. Husaini, M.; Kamarudin, L.; Zakaria, A.; Kamarudin, I.K.; Ibrahim, M.; Nishizaki, H.; Toyoura, M.; Mao, X. Non-Contact Breathing Monitoring Using Sleep Breathing Detection Algorithm (SBDA) Based on UWB Radar Sensors. *Sensors* **2022**, *22*, 5249. [[CrossRef](#)]
57. Analia, R.; Forster, A.; Xie, S.-Q.; Zhang, Z. Privacy-Preserving Approach for Early Detection of Long-Lie Incidents: A Pilot Study with Healthy Subjects. *Sensors* **2025**, *25*, 3836. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.