Video-Based Elevated Skin Temperature Detection

Ananyananda Dasari ⬤, Ambareesh Revanur, Laszlo A. Jeni ⬤, and Conrad S. Tucker ⬤

Abstract—**In this work, we propose a non-contact video-based approach that detects when an individual's skin temperature is elevated beyond the normal range. The detection of elevated skin temperature is critical as a diagnostic tool to infer the presence of an infection or an abnormal health condition. Detection of elevated skin temperature is typically achieved using contact thermometers or non-contact infrared-based sensors. The ubiquity of video data acquisition devices such as mobile phones and computers motivates the development of a binary classification approach, the Video-based TEMPerature (V-TEMP) to classify subjects with non-elevated/elevated skin temperature. We leverage the correlation between the skin temperature and the angular reflectance distribution of light, to empirically differentiate between skin at non-elevated temperature and skin at elevated temperature. We demonstrate the uniqueness of this correlation by 1) revealing the existence of a difference in the angular reflectance distribution of light from skin-like and non-skin like material and 2) exploring the consistency of the angular reflectance distribution of light in materials exhibiting optical properties similar to human skin. Finally, we demonstrate the robustness of V-TEMP by evaluating the efficacy of elevated skin temperature detection on subject videos recorded in 1) laboratory controlled environments and 2) outside-the-lab environments. V-TEMP is beneficial in two ways; 1) it is non-contact-based, reducing the possibility of infection due to contact and 2) it is scalable, given the ubiquity of video-recording devices.**

Index Terms—**Angular reflectance, non-contact, skin temperature, video-based.**

I. INTRODUCTION

THERMO-regulatory control mechanisms are vital to the maintenance of thermal equilibrium within the human body [1]. Deviations in thermal equilibrium conditions due to diseases such as COVID-19 or breakdown of the thermo-regulatory mechanism due to environmental causes such as heat stroke, manifest as elevated body temperature [2], leading to a corresponding rise in skin temperature. Detection of elevated body or skin temperature is critical as a diagnostic screening tool for various medical ailments [3], [4], [5]. The core body temperature is defined as the temperature of the blood in the pulmonary artery, which can be measured by invasive methods such as the Pulmonary Artery Thermistor (PAT) [6]. The rectal temperature is accepted as the gold standard measurement for the core body temperature [7]. However, rectal temperature measurement carries the risk of infection and is not implementable in non-clinical settings [7], [8]. Contact-based sensors such as mercury-based clinical thermometers and digital thermometers provide an approximation to the core body temperature and can be utilized in non-clinical settings. However, contact-based sensors, similar to invasive methods, carry the risk of transmitting infections between subjects, which necessitates additional sanitizing procedures [9], [10]. In order to mitigate the risk of infection and also reduce measurement/sanitizing times, the skin temperature can be detected as an alternative to the core body temperature. The skin is the interface between the body and the environment, which makes it readily accessible for temperature measurement as compared to sites of core body temperature measurement such as the ear canal or the oral cavity [11].

Existing methods of detecting elevated skin temperature can be classified into contact-based methods and non-contact-based [12], [13] methods. Contact-based approaches for elevated skin temperature detection utilize photoplethysmography (PPG) [14] techniques to monitor changes in blood flow patterns under the skin to determine the amplitude of skin temperature oscillation. [14]. PPG approaches can measure the amplitude of skin temperature oscillation given an initial temperature, by utilizing an empirically determined correlation between skin blood flow and oscillations in the temperature signal [15]. However, due to the low oscillation frequency of the temperature signal, PPG sensors cannot detect elevated skin temperature as detection of elevated skin temperature requires knowledge of the absolute skin temperature [16], [17], [18]. Other contact-based approaches [19] measure optical and thermal properties of the skin (such as skin emissivity) to obtain a measure of the skin temperature. The multilayered composition and heterogeneous color distribution of the skin create difficulties in measuring the optical and thermal properties of the skin, which limits the use of contact-based sensors in real-time applications. Non-contact skin temperature estimation approaches such as infrared sensors have shown promise for deployment in real-world environments [20], [21]. However, the ubiquity of devices such as computers and mobile phones that can record Red-Green-Blue

Ananyananda Dasari is with the Mechanical Engineering Department, Carnegie Mellon University, USA.

Ambareesh Revanur and Laszlo A. Jeni are with the Robotics Institute, Carnegie Mellon University, USA.

Conrad S. Tucker is with the Mechanical Engineering Department, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: conradt@andrew.cmu.edu).
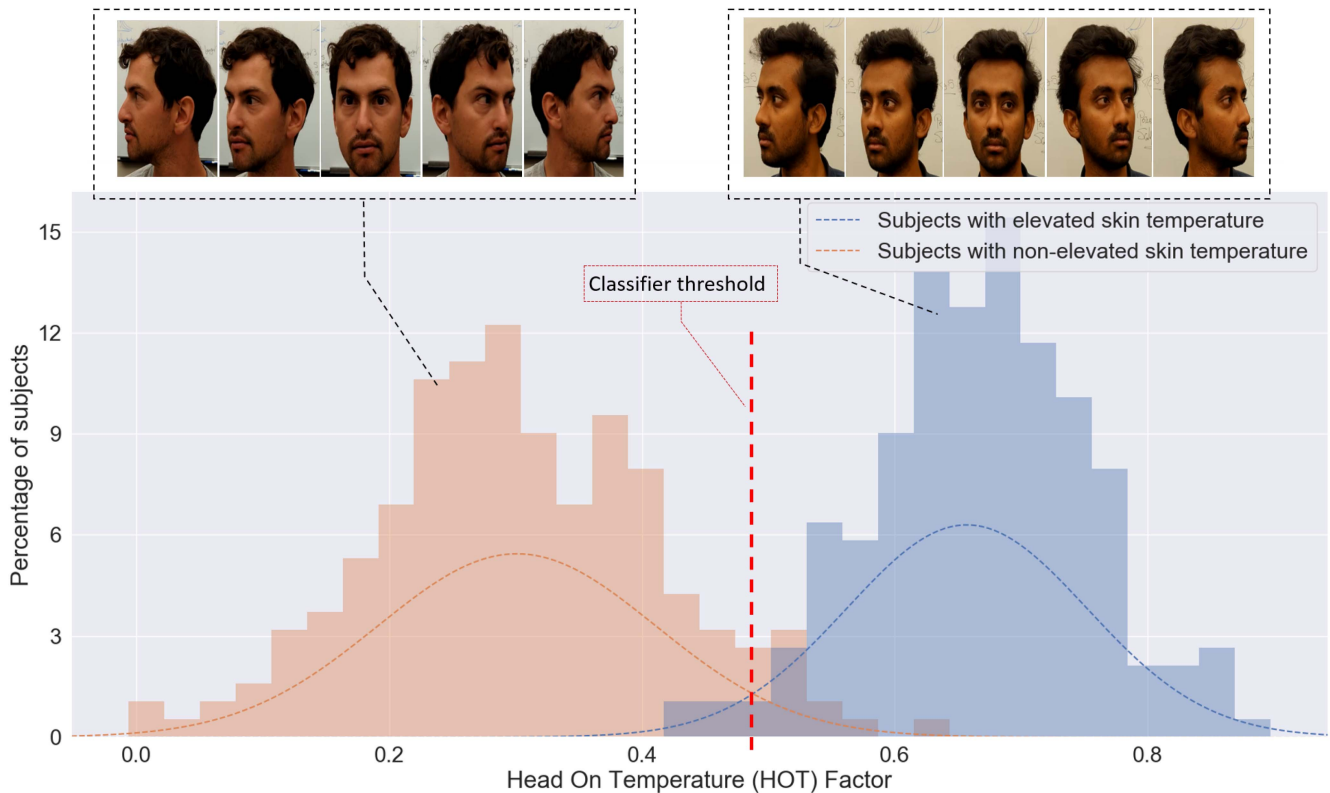
Fig. 1. Determination of the Head On Temperature factor (*HOT factor*). Representative figure showing the determination of the *HOT factor* (an intermediary parameter) from a facial video to act as a substitute for the actual temperature of a person. Subjects with elevated skin temperature are detected using a threshold classifier based on the *HOT factor*, over multiple iterations. Written informed consent was obtained from the subjects to display their identifiable facial images.

(RGB) video motivates the development of a non-contact, non-infrared approach to detect elevated skin temperature.

We propose V-TEMP, a binary classification approach to classify subjects into two categories: subjects with an elevated skin temperature and subjects with a normal (non-elevated) skin temperature. Elevated skin temperature typically indicates fever and other conditions such as heatstroke [22], [23], [24], [25], [26], [27], [28], [29]. Video data of a person's face is utilized in this work, instead of other regions of the body (such as the fingers) as the face region shows a steady increase of skin temperature in response to an increase in core body temperature [28]. In contrast, fever typically manifests itself as oscillating skin temperature in the extremities of the body, such as the fingers [26], even causing a decrease of finger skin temperature due to vasoconstriction in some cases [23], [26], [27], [28].

This work presents V-TEMP for detecting elevated skin temperature using a standard video with a minimum resolution of 320×240 pixels, recorded at a minimum of 30 frames per second (fps) with a camera found in devices such as computers and mobile phones. Our approach is beneficial in two ways; (1) it is non-contact-based, reducing the possibility of infection due to contact, (2) it is scalable, given the ubiquity of video-recording devices. The core principle behind V-TEMP is that a change in camera viewing angle causes a change in the amount of light reflected by the skin, due to the angular variation of skin properties such as skin reflectance. Studies have theoretically

and experimentally explored the principles of optical scattering from human skin at different wavelengths [30], [31], [32], [33], [34]. Comparative studies on multi-angular, multi-spectral optical scattering have been performed for different materials highlighting the angular variation of optical properties in different materials as compared to that of human skin [35], [36], [37], [38], [39]. Investigating the uniqueness of the angular skin reflectance distribution, we empirically find that the optical properties of the skin and skin-like materials differ significantly from optical properties of non-living objects. Also, the optical properties of skin differ significantly from the optical properties of non-skin regions of the human face. We utilize the aforementioned unique difference in skin optical properties to infer the presence of a temperature-like physiological signal from the facial skin. Additionally, we empirically find that the optical properties of the skin differ for skin at non-elevated temperatures as compared to skin at elevated temperatures (Fig. 1) for different camera viewing angles. We utilize this difference to empirically develop a threshold classifier that can classify a subject as having normal (non-elevated) or elevated skin temperature.

Research on the detection of elevated skin temperature has focused on estimating the absolute skin temperature and identifying anomalies. Prior work on skin temperature estimation methods can be classified into contact-based and non-contact-based approaches. The following sections review the contact-based and the non-contact-based skin temperature estimation approaches, highlighting the contributions of the current work.

A. Contact-Based Skin Temperature Estimation

Researchers have explored contact-PPG-based skin temperature estimation using a pulse-oximeter by establishing a relationship between human physiology and skin temperature. Sun and Thakor provide an analysis of the physiological information that can be extracted from the different frequency components of the PPG signal [40]. The PPG signal comprises of five different frequency components ranging from 0.007 Hz to 5 Hz, corresponding to respiration, blood pressure, autonomous nervous system, thermo-regulation and pulse signal respectively [14], [41]. Unfortunately, the pulse rate (high frequency component) and the average blood volume/respiration signal (low frequency component) dominate the PPG signal [14]. Further, the thermo-regulatory signal composes part of the low frequency component in the range of 0.0039–0.05 Hz [18], [42]. The low frequency range of the thermo-regulatory signal necessitates a high resolution sensor and longer measurement time [16], [17], [18]. Jeong et al. attempted to amplify the thermo-regulatory signal frequency by subjecting the arm to heating/cooling using a water bath [43]. The requirement of heating/cooling of the measuring area requires a water bath or cold compress and limits the use of this method for temperature estimation in an outside-the-lab environment. The study also empirically determined a correlation between cardiac parameters (such as pulse rate, stroke volume and cardiac output) and changes in skin temperature. Sanchez et al. induced changes in skin emissivity by regulating external air temperature and established a correlation between skin emissivity and skin temperature [19]. Such regulation of air temperature requires a carefully-controlled experimental setup and is not suitable for deployment in outside-the-lab settings.

In summary, contact-based approaches pose a risk of infection especially in cases involving a risk of communicable disease spread such as COVID-19. This motivates the advancement of non-contact-based approaches which will be explained in detail in the following section.

B. Non-Contact-Based Temperature Estimation

Non-contact-based approaches such as infrared sensor-based approaches have been utilized to estimate the skin temperature of a person in real time, and have the advantage of reducing the risk of infection due to the lack of physical contact between the skin and the sensor [44]. Infrared sensors detect the amount of light energy radiated by an object and determine the temperature of the object using thermodynamic and signal processing techniques [45], [46], [47], [48]. The approaches advanced by Cho et al. and Kocoglu et al. measure the temperature of the tympanic membrane (inner ear) that can be utilized as a surrogate for the core body temperature [49], [50]. The developed approaches require frequent sensor adjustment to fit the shape of the ear canal. Ng et al. have developed an infrared sensor capable of using the forehead as a measurement site [51]. Such a method lacks physical contact, minimizing the risk of infection. A study by Chaglla et al. developed a sensor for continuous monitoring of ear temperature using a graphene coating on the lens of the infrared thermopile sensor [52]. The method was validated with clinical thermometer readings. Sharma and Yadav [53] proposed a temperature estimation algorithm using a video feed from an infrared camera and detect facial regions using the Viola Jones [54] algorithm. The results of the method were validated using a digital thermometer. The study, however, requires video feed from an infrared camera, limiting its scalability. Rodriguez-Lozano [12] et al. proposed a segmentation method to identify the forehead and calculate the temperature using the image from an infrared camera. The study also analyzes the performance of infrared sensing approaches in cases such as people with facial hair, glasses and exhibiting neck rotations. A study by Chin et al. developed a modified infrared camera model to enable mass screening of people in crowded settings [55]. Dell'Isola et al. developed a screening rule to detect subjects with elevated temperature using an infrared camera [11].

In this work, we utilize an RGB camera instead of an infrared camera, which allows the integration of the method with a smartphone application or other camera-based systems such as digital cameras, thereby increasing its potential scale and accessibility due to the ubiquity of video-recording devices [56], [57]. Studies by [61], [65] found that the measured skin temperature drops by more than 4° F with an angular deviation of 45° and above from normal incidence. Arieli et al. have empirically studied the differences in the magnitude of apparent skin temperature drop with a change in infrared camera viewing angle for different skin temperatures [66]. The study also found that the apparent drop in skin temperature with increasing viewing angle is steeper for higher skin temperatures. However, the study does not develop a method to detect elevated skin temperature using the change of apparent skin temperature with camera viewing angle. In order to leverage the dependence of the change in apparent skin temperature with camera angle using a facial video, we calculate the skin reflectance at different instances, corresponding to different camera angles. Measuring the absolute skin temperature using the calculated skin reflectance requires knowledge of ambient temperature and skin-specific parameters such as emissivity, which are difficult to measure using a facial video. To circumvent the need to measure skin-specific and atmospheric properties, we calculate an intermediate ratio called the 'Radiatively Inspired Reflectance ratio' (RIR ratio) instead of attempting to calculate the absolute temperature. We utilize the RIR ratio to denote the variation of apparent temperature with camera pose and calculate a *'Head On Temperature factor (HOT factor)'* to classify subjects as having a normal (non-elevated) or elevated skin temperature, based on a binary threshold classifier - V-TEMP. To demonstrate the determination of a unique *HOT factor* from facial videos, we utilize different datasets consisting of objects with different reflectance distributions and find that we can determine a *HOT factor* only for videos with a reflectance distribution similar to real facial skin. The determination of the *HOT factor* is critical for V-TEMP to make an inference about the elevated/non-elevated nature of the temperature of the skin. Table I shows a summary of the existing works in the area of skin temperature estimation and anomalous temperature detection.

TABLE I
SUMMARY OF EXISTING HUMAN BODY TEMPERATURE ESTIMATION METHODS

| Authors	Year	Contact-based sensor	Non-contact infrared	Non-contact, non-infrared	Lab Control Environment	Outside-the-lab study	Elevated Skin Temperature Detection
Betta et al. [45]	1997		✓		✓		
Valle et al. [9]	1999	✓	✓		✓		
Craig et al. [8]	2000	✓			✓		
Craig et al. [10]	2002	✓	✓		✓		
Kocoglu et al. [50]	2002		✓		✓		
Pusnik et al. [47]	2004		✓				
Ng et al. [51]	2005		✓		✓		
Sanchez-Marin et al. [19]	2009	✓			✓		
Sagaidachnyi et al. [17]	2012	✓			✓		
Apa et al. [74]	2013	✓	✓		✓		
Sagaidachnyi et al. [18]	2014	✓			✓		
Jeong et al. [43]	2014	✓			✓		
Tay et al. [75]	2015		✓		✓		
Sharma and Yadav [53]	2017		✓		✓		
Arieli et al. [66]	2017		✓		✓		
Chaglla et al. [52]	2018		✓		✓		
Rodriguez-Lozano et al. [12]	2019		✓		✓		
Hsiao et al. [76]	2020	✓	✓				
Khan et al. [44]	2021	✓	✓		✓		
Shajkofci et al. [48]	2021		✓		✓		
Cho et al. [49]	2021	✓	✓		✓		
Chin et al. [55]	2021		✓		✓	✓	
Dell'Isola et al. [11]	2021		✓				
Khaksari et al. [77]	2021		✓				
Current Study (V-TEMP)	2021			✓	✓	✓	✓
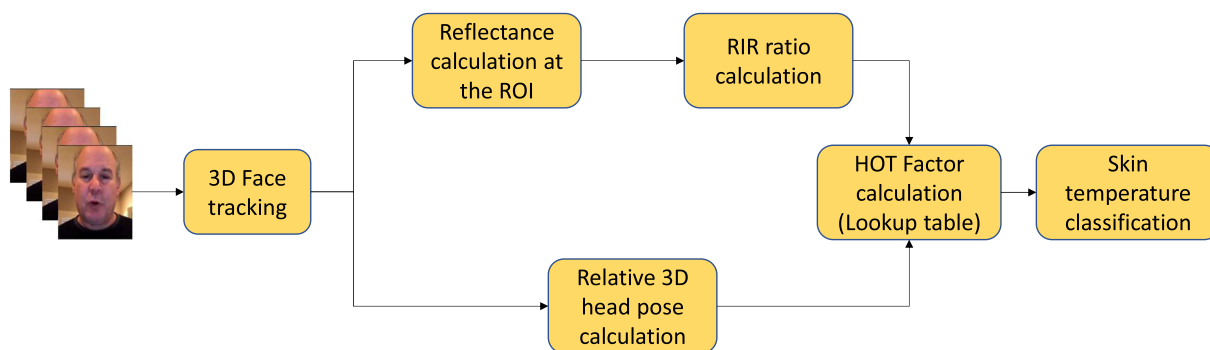
Fig. 2. Schematic for V-TEMP. Schematic diagram of V-TEMP for detecting elevated skin temperature of a person from a facial video, demonstrating steps such as the determination of the RIR ratio and the *HOT factor*. The depiction of the subject's facial image is in accordance with the Fair Use Policy of the video sharing platform [90].

V-TEMP is tested on standard video datasets consisting of healthy individuals [67], [68] and also on a dataset of subject videos in both laboratory conditions as well as diverse real-world environments. Testing on a dataset with subjects in outside-the-lab conditions is essential to evaluate the performance and deployability of non-contact vital sign monitoring algorithms. The study by Dasari et al. demonstrates the high variability of state-of-the-art remote photoplethysmography methods when deployed in outside-the-lab conditions, as compared to laboratory settings, highlighting the need for evaluating V-TEMP on subjects in outside-the-lab settings [73]. The performance of V-TEMP in diverse real-world conditions is found to be comparable to the performance in laboratory settings, highlighting its promise for real-world deployment.

II. METHODS

A. Overview

Fig. 2 shows the flowchart of V-TEMP. The study by Arieli et al. [66] describes the variation of measured skin temperature with a change in camera viewing angle for a ground truth skin temperature defined at normal incidence. To reduce the
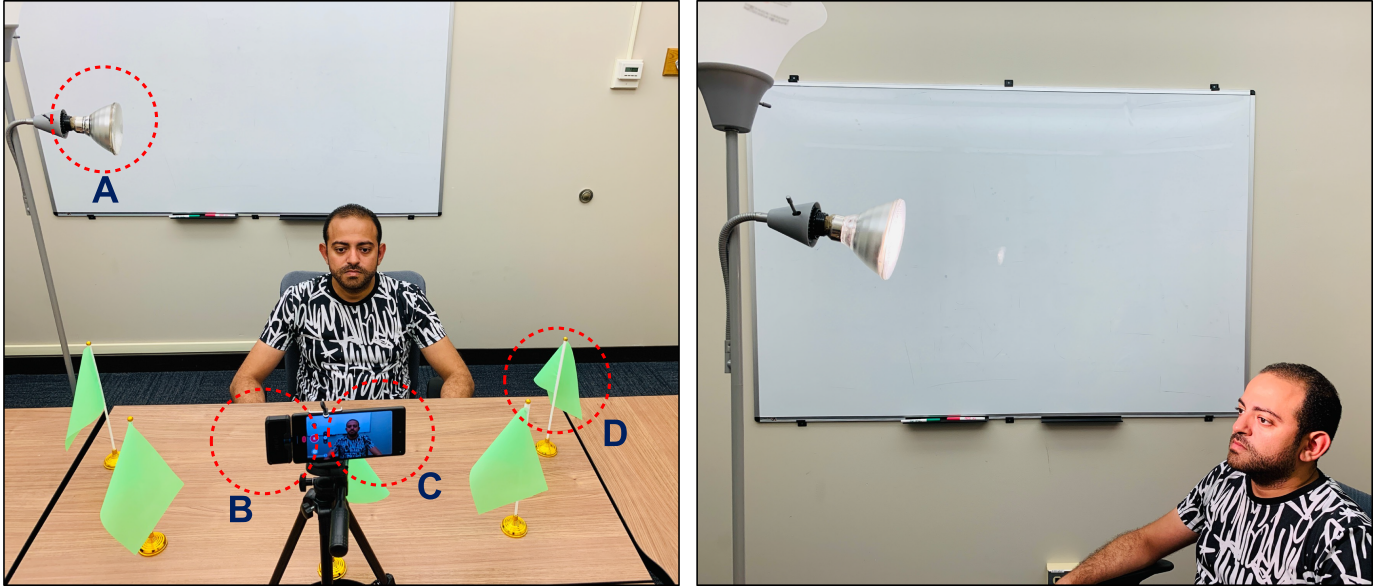
Fig. 3. NEST-Lab data collection setup. Experimental setup for collecting subject videos for the NEST-Lab dataset showing overall setup for video recording (left) with A - Heat lamp, B - Thermal camera, C - Mobile phone and D - Guiding flags. Artificial skin heating in progress with the heat lamp focused normally on the subject's face (right). Written informed consent was obtained from the subject to display their identifiable facial image.

dependence on the ground truth skin temperature at normal incidence, we introduce a 'Radiatively Inspired Reflectance ratio (RIR ratio)' to denote the variation of apparent skin temperature at different viewing angles, as compared to normal incidence. We utilize the calculated RIR ratio and the camera viewing angle to determine the *HOT factor* from the lookup function we generated from the results of the study by [66], which documented the changes in apparent skin temperature with the camera viewing angle. To determine the RIR ratio and the camera viewing angle at each frame of the video, we need to track the subject's face throughout the video using a face tracking algorithm. We employ a real-time face tracking algorithm called ZFace for obtaining facial regions-of-interest (ROI). ZFace is a real-time 3D cascade regression-based approach developed for tracking facial landmarks. ZFace is robust to illumination conditions and does not make any assumptions about skin properties. Therefore, ZFace is suitable for tracking the facial landmarks for skin-temperature classification. Along with ROIs, the ZFace tracker also provides 512 two-dimensional landmarks and the head pose in terms of the fundamental rotation angles (yaw, pitch and roll) for each frame of the video (see Fig. 3 for an example). We determined empirically that landmarks present on the nose region/between the eyes are valuable for skin-temperature classification (Section: *Experimental results*). Further, the nose/between the eyes landmarks are also completely visible during profile-to-profile movement of the head.

After extracting the facial ROIs, we obtain the reflectance of the ROI by computing the mean pixel intensity. Due to the directional dependence of skin properties, the calculated value of the reflectance is modified using the 'Bidirectional Reflectance Function' (BDRF) [85], [86], which explains the relationship between intrinsic skin reflectivity changes and camera viewing angle (or head pose). Accordingly, the RIR

ratio ($\frac{T_i}{T_j}$) with the first frame of the video is calculated using the Temperature-Reflectance model developed (Section: *Temperature-Reflectance model*). The RIR ratio is calculated as

$$\frac{T_i}{T_j} \approx \left[\frac{\tau_{atm} \left(I_{p_i} - 1 \right) + 1}{\tau_{atm} \left(I_{p_j} - 1 \right) + 1} \right]^{1/4} \qquad (1)$$

where T_i and T_j are the temperatures of the skin at two different head poses (i and j), τ_{atm} is the transmittivity of the atmosphere, I_{p_i} and I_{p_j} are the pixel intensities of the tracked facial landmark at the two head poses i and j. Head pose i is the base head pose (defined by the head pose in the first frame of the video), where the head is aligned with the camera line of sight. Head pose j is the head pose at any instant in the video. The difference in the yaw angle with respect to the first frame of the video provides a measure of subject head angular deviation for the particular frame ($|yaw_i - yaw_j|$). The RIR ratio which is determined using the Temperature-Reflectance model and the subject head angular deviation are provided as input to the lookup function which outputs the *HOT factor* value. Fig. 4 shows the lookup surface obtained from the lookup function. The RIR ratio calculated from the Temperature-Reflectance model is shown on the z-axis with the subject head angular deviation shown on the x-axis. The y-axis value obtained for a particular temperature ratio and angular deviation denotes the *HOT factor* which is utilized to classify the subject as having normal or elevated skin temperature using a threshold classifier.

V-TEMP depends on various parameters such as the chosen face region of interest, the size of the face region of interest, and the threshold level of the threshold classifier. Experiments are conducted to identify the optimal parameters for temperature estimation and to evaluate the robustness of the approach, by testing across varying datasets as described in Section: *Datasets*.
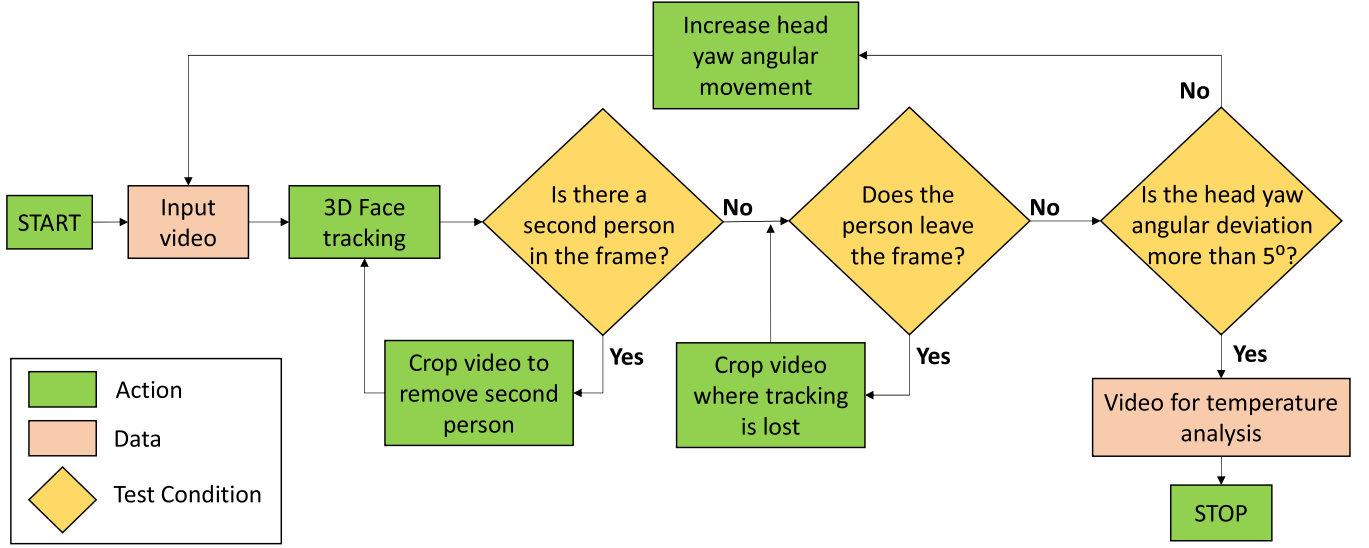
Fig. 4.   Process flow for selecting a video for analysis. Process flow chart to record a video suitable for temperature analysis demonstrating processes such as face tracking, video cropping and removal of multiple people in the video frame.
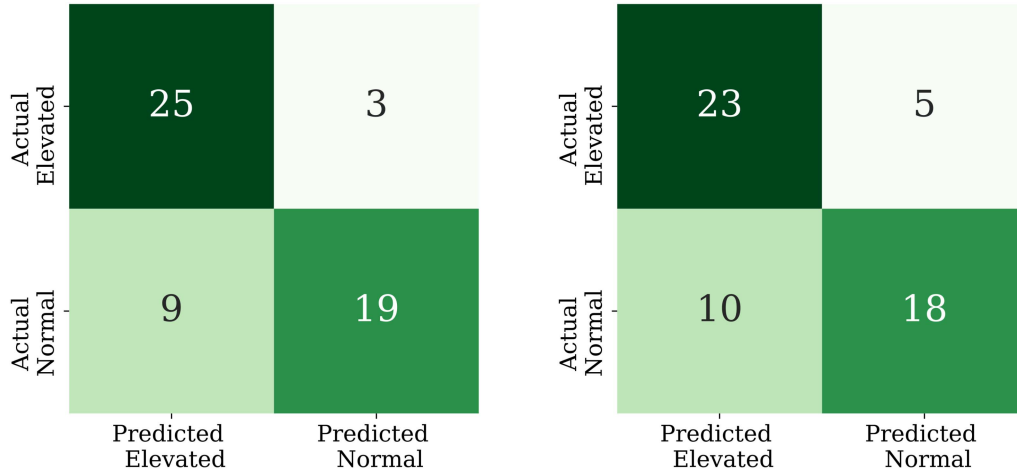

Fig. 5.   Experimental results for V-TEMP in laboratory conditions and outside-the-lab conditions. Confusion matrices for V-TEMP, evaluated on videos from the NEST-Lab dataset (left) and the NEST-Wild dataset (right).

B. Temperature-Reflectance Model

We develop a relation between the temperature and reflectivity of the skin using the thermal model of the infrared camera (Fig. 5). Every object at a temperature above absolute zero (0 Kelvin) radiates energy in the form of electromagnetic waves [58]. According to the Stefan-Boltzmann law [59], [60], the radiation emitted (W) by a surface at a temperature T Kelvin, is given by

$$W = \epsilon \sigma T^4 \qquad (2)$$

where ϵ is the emissivity of the body, and σ is the Stefan-Boltzmann constant. The general model for temperature measurement with an infrared camera is presented in Fig. 5. The total energy received by the infrared camera (W_{tot}) is given by [61], [62]:

$$W_{tot} = W_{obj} + W_{refl} + W_{atm} \qquad (3)$$

where W_{obj} is the energy emitted by the body, W_{refl} is the energy emitted by the surroundings, reflected by the object and W_{atm} is the energy emitted by the atmosphere. Using the Stefan Boltzmann law, W_{obj} can be written as

$$W_{obj} = \epsilon_{obj} \sigma T_{obj}^4 \qquad (4)$$

where ϵ_{obj} is the emissivity of the object, σ is the Stefan-Boltzmann constant, T_{obj} is the temperature of the object. Since part of the radiation from the object is absorbed by the atmosphere (Fig. 5), W_{obj} is multiplied by the atmospheric transmittivity τ_{atm}. This gives the following expression for W_{obj}.

$$W_{obj} = \epsilon_{obj} \sigma \tau_{atm} T_{obj}^4 \qquad (5)$$

W_{refl} is the radiation from the atmosphere, reflected from the object and partly re-absorbed by the atmosphere. W_{refl} can be

written as

$$W_{refl} = \rho_{obj}\sigma\tau_{atm}T_{refl}^4 \tag{6}$$

where ρ_{obj} is the reflectivity of the object and T_{refl} is the ambient temperature. For the current application, the ambient temperature can be approximated by the atmospheric temperature (T_{atm}).

$$\Rightarrow T_{refl} \approx T_{atm} \tag{7}$$

This gives the expression for W_{refl} as

$$W_{refl} = \rho_{obj}\sigma\tau_{atm}T_{atm}^4 \tag{8}$$

W_{atm} is the radiation emitted by the atmosphere and directly reaching the camera. W_{atm} can be expressed as

$$W_{atm} = \epsilon_{atm}\sigma T_{atm}^4 \tag{9}$$

Combining the expressions for W_{obj}, W_{refl} and W_{atm}, we obtain the following expression for W_{tot}.

$$W_{tot} = \epsilon_{obj}\sigma\tau_{atm}T_{obj}^4 + \rho_{obj}\tau_{atm}\sigma T_{atm}^4 + \epsilon_{atm}\sigma T_{atm}^4 \tag{10}$$

This gives

$$T_{obj} = \left(\frac{W_{tot} - \rho_{obj}\tau_{atm}\sigma T_{atm}^4 - \epsilon_{atm}\sigma T_{atm}^4}{\epsilon_{obj}\sigma\tau_{atm}}\right)^{\frac{1}{4}} \tag{11}$$

The temperature T_{obj} recorded by the infrared camera is the apparent temperature of the object. According to Stefan-Boltzmann law, the energy radiated by the object due to its apparent temperature is given by

$$W_{app} = \epsilon\sigma T_{obj}^4 \tag{12}$$

W_{app} is the energy that would have been recorded by the infrared camera in the absence of atmospheric influence. Due to the influence of the atmosphere, the infrared camera records a different amount of radiant energy from the object, given by W_{obj}. Accounting for the contribution of atmospheric radiation, the total radiant energy received by the infrared camera is given by W_{tot}. Through a generic temperature calibration, the infrared camera attempts to minimize the difference between W_{app} and W_{tot}, providing us with the following approximation.

$$W_{app} \approx W_{tot} \tag{13}$$

Such calibration is performed by an infrared camera to relate the infrared radiation received to the actual temperature of the object. We note that an RGB camera performs a deterministic modification of the visible light intensities received and does not relate the intensities to an external entity such as the temperature. Expanding (12) using (9) and (11), we obtain

$$\epsilon\sigma T_{obj}^4 \approx \epsilon_{obj}\sigma\tau_{atm}T_{obj}^4 + \rho_{obj}\tau_{atm}\sigma T_{atm}^4 + \epsilon_{atm}\sigma T_{atm}^4 \tag{14}$$

where, ϵ_{obj} and ρ_{obj} are the emissivity and reflectivity of the body and ϵ_{atm} and τ_{atm} are the emissivity and transmittance of the atmosphere. Atmospheric emissivity (ϵ_{atm}) can be expressed as $\epsilon_{atm} = 1 - \tau_{atm}$ [62]. Substituting in (14) and rearranging, we obtain

$$T_{obj}^4 \approx \frac{(\tau_{atm}(\rho_{obj} - 1) + 1)T_{atm}^4}{\epsilon_{obj}(1 - \tau_{atm})} \tag{15}$$

Equation (15) gives the temperature at one instance of the video corresponding to one camera (or head) pose. We consider two instants of the video i and j, when the camera pose is given by $pose_i$ and $pose_j$, as the reflectance is observed to change with viewing angle. We denote T_i and T_j as the corresponding temperatures of the object at these instants. This gives us the following relations for T_i and T_j. The reflectance of the object at the two instances is different, given by ρ_{obj_i} and ρ_{obj_j} respectively.

$$T_i^4 \approx \frac{(\tau_{atm}(\rho_{obj_i} - 1) + 1)T_{atm}^4}{\epsilon_{obj}(1 - \tau_{atm})} \tag{16}$$

$$T_j^4 \approx \frac{(\tau_{atm}(\rho_{obj_j} - 1) + 1)T_{atm}^4}{\epsilon_{obj}(1 - \tau_{atm})} \tag{17}$$

Dividing (16) and (17) to eliminate the atmospheric temperature T_{atm}, we obtain

$$\frac{T_i}{T_j} \approx \left[\frac{\tau_{atm}(\rho_{obj_i} - 1) + 1}{\tau_{atm}(\rho_{obj_j} - 1) + 1}\right]^{1/4} \tag{18}$$

Considering two different instants of the video, i and j also ensures that any calibration performed within the RGB camera does not affect the outcome of (18), as far as the calibration is consistent throughout the video. Hence, the calibration, if any, that is present in the RGB camera should not affect the performance of the algorithm.

Equation (18) gives the relation between the ratio of apparent temperatures of an object at two different camera poses and the thermal reflectance of the object at those poses. We approximate the thermal reflectance term (ρ_{obj}) by the visible reflectance term ($\rho_{v_{obj}}$) to enable the calculation of the temperature ratio from an RGB video, as given by (18). Such a substitution of thermal reflectance with visible reflectance leverages the correlation between changes in thermal and visible reflectance with camera pose (Supplementary Information).

Since reflectance is defined as the ratio of the magnitudes of reflected light and incident light on a surface, $\rho_{v_{obj}}$ can be expressed in terms of the normalized pixel intensities as

$$\rho_{v_{obj}} = I_p \tag{19}$$

where I_p is the normalized intensity of the selected pixel [63], [64]. Min-max normalization is performed for each video. The normalized pixel intensity I_p is obtained from the mean normalized pixel intensities of a region of interest.

$$\frac{T_i}{T_j} \approx \left[\frac{\tau_{atm}(\rho_{v_{obj_i}} - 1) + 1}{\tau_{atm}(\rho_{v_{obj_j}} - 1) + 1}\right]^{1/4} \tag{20}$$

Rewriting (20) using the expression for $\rho_{v_{obj}}$, (19) we obtain the RIR ratio, $\frac{T_i}{T_j}$ as

$$\frac{T_i}{T_j} \approx \left[\frac{\tau_{atm}(I_{p_i} - 1) + 1}{\tau_{atm}(I_{p_j} - 1) + 1}\right]^{1/4} \tag{21}$$

The normalized pixel intensities I_{p_i} and I_{p_j} are determined from the video using the identified face region of interest. We note that the contribution of the atmospheric radiation provides the constant terms in (21), when we utilize the ratio of the apparent
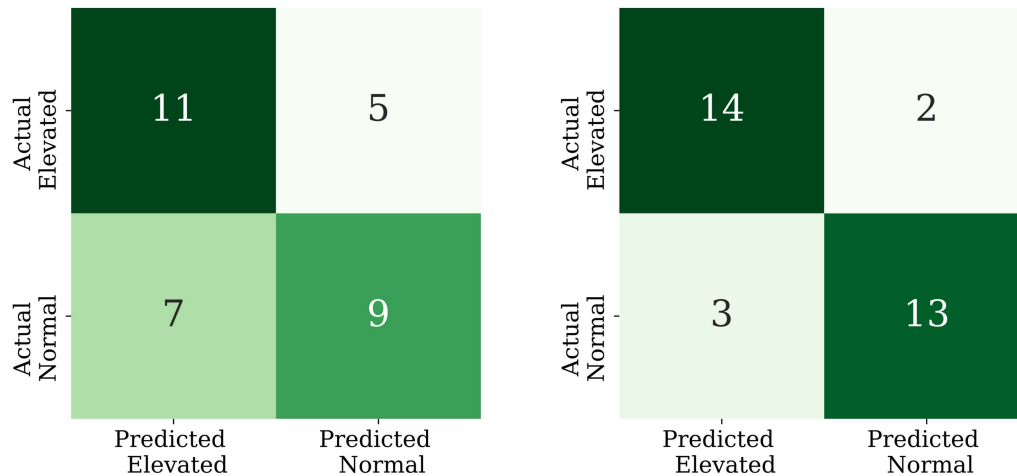
Fig. 6. Experimental results for V-TEMP in laboratory conditions. Confusion matrices for videos sourced from the NEST-Lab dataset with lesser than 5° subject head angular deviation (left) and greater than 25° subject head angular deviation (right).

temperature at two instants of the video (i and j). Using such a ratio-based method, the atmospheric contribution in (14) is reduced into the constant terms (15) and (18) and hence will not impact the RIR ratio calculation since the pixel intensities undergo a min-max normalization (which ensures each video has the same range of pixel intensities). Introducing atmospheric contribution into a non-ratio-based (15) can cause significant impact, both when the pixel values are normalized as well as non-normalized. However, in this approach, we are using a ratio-based method and hence the constant terms appearing due to atmospheric radiation do not impact the RIR ratio if the pixel values are normalized. We note that using a different normalization procedure will change the range of pixel intensities, affecting the threshold for binary classification. The threshold will need to be recomputed based on the new pixel intensity range as described in Section: Experimental Results

The obtained RIR ratio $\frac{T_i}{T_j}$ is input to the lookup function along with the subject head angular deviation between the ith and j^{th} instances. The lookup function outputs the *HOT factor* that can be utilized to classify the subject as having elevated or non-elevated skin temperature using a threshold classifier. For objects other than human faces, the reflectance can be extracted by defining a similar region of interest by tracking the object through the duration of the video. This extracted reflectance can be utilized to calculate the RIR ratio using (21). The calculated RIR ratio is utilized to obtain a *HOT factor* using a lookup function, specific to the object. The lookup function for human skin is generated based on results from previous studies [66] which study the variation of the apparent skin temperature with change in yaw angle at different base skin temperatures. The *HOT factor* (having a range of 0 – 1) is a substitute for the actual skin temperature and provides the link between the calculated RIR ratio and the subject head angular deviation. The *HOT factor* is the output of the lookup function and is further utilized for threshold-based binary classification of a subject video as having elevated/non-elevated skin temperature. We have modified the previous lookup function [66] to denote the change in the RIR ratio with yaw angle instead of the change in the apparent skin

temperature with the yaw angle (Section: Non-Contact-Based Temperature Estimation). Since the change in apparent skin temperature with yaw angle is different at different base skin temperatures, the change in RIR ratio with yaw angle also depends on the base skin temperature. However, since we are now using the RIR ratio instead of the apparent skin temperature, we introduce a different factor instead of the base skin temperature. To this end, we define the *HOT factor* as the differentiating factor between the RIR ratio-yaw angle curves. Since the *HOT factor* is a substitute for the base skin temperature, we can now utilize it to define the threshold to differentiate between non-elevated and elevated skin temperature, based on experimentation. Based on our experiments, the *HOT factor* range of 0–1 maps to a base skin temperature range of 295 K to 305 K, which is the skin temperature range (of subjects with normal skin temperature) we observed in our experiments and was also reported in existing studies [66].

C. Experimental Protocol to Develop the NEST-Lab Dataset

The experimental setup (Fig. 3) consists of a mobile phone (Google Pixel 6) mounted on a tripod stand in the landscape orientation. The landscape orientation allows for a larger field of view to be captured. The RGB video is recorded at a frame rate of 60 fps and a resolution of 1920 × 1080 pixels. The FLIR One Pro thermal camera is connected to the Google Pixel 6 mobile phone and records the thermal video (ground truth for skin temperature) simultaneously with the RGB video. The thermal video is recorded at a frame rate of 9 fps and a resolution of 1440 × 1080 pixels. The Pixel-FLIR setup (Fig. 6) is placed at a distance of 0.7 m from the seated subject. The heat lamp is placed at a distance of 0.5 m [87], [88] from the subject with the lamp focused normally on the subject's face, to ensure uniform heating of all facial regions. The subject is instructed to perform left-right head rotation for the duration of the video recording, guided by 5 flags placed on the table at angles of 30°, 60°, 90°, 120° and 150° from left to right. Such head rotation provides

TABLE II
STATISTICS OF DIFFERENT DATASETS UTILIZED FOR THE STUDY

Dataset	Number of videos	Video contents	Videos with normal skin temperature	Videos with elevated skin temperature	Head movement
Objectron	90	Non-living objects	NA	NA	$35.1° \pm 15.2°$
BP4D+	32	Real human faces	32	0	$2.7° \pm 3.5°$
3DFAW	33	Real human faces	33	0	$22.7° \pm 20.8°$
PSU-HR	24	Real human faces	24	0	$1.8° \pm 0.6°$
NEST-Lab	56	Real human faces	28	28	$34.3° \pm 5.1°$
NEST-Wild	56	Real human faces	28	28	$14.9° \pm 15.3°$

the subject head angular deviation required for the V-TEMP algorithm. Informed consent is obtained from the subject in accordance with the relevant guidelines and regulations of The Carnegie Mellon University's *Institutional Review Board* (IRB).

The experimental protocol to record subject videos for the NEST-Lab dataset consists of 1) ensure subject is seated upright facing the Pixel-FLIR setup 2) record RGB and thermal video of the subject's face simultaneously for a duration of 120 seconds while the subject makes left-right head motions 3) shine the Philips IR heat lamp on the subject's face for a duration of 180 seconds for artificial skin temperature elevation 4) record RGB and thermal video of the subject's face simultaneously for a duration of 120 seconds while the subject makes left-right head motions.

Artificial facial skin temperature elevation is achieved by exposure to an infrared heat lamp (Philips PAR38E - 175W [89]). Studies by Kim et al. [87] and Lopes et al. [88] indicated the rate of elevation in skin temperature to be approximately $2°F$ per minute of head-on exposure to the heat lamp. To cause an elevation in skin temperature similar to a fever, an exposure time of 3 minutes was determined to be optimal to avoid discomfort to the subject and to allow for sufficient ($5°F$) elevation in skin temperature (Figs. 4 and 5). We collected videos from 32 subjects in a temperature-controlled ($74°F$) laboratory setting with a white background and constant ambient lighting. The recorded videos are collected into the NEST-Lab dataset and are utilized to evaluate the performance of V-TEMP as described in Section: Datasets. The ground truth skin temperature obtained using the thermal camera is considered when determining the skin temperature status (elevated/non-elevated) of a subject. Additional information is provided in the supplementary information.

III. RESULTS AND DISCUSSION

A. Datasets

The datasets utilized in this study differ in terms of the number of videos, the contents of the videos (real human faces, synthetic faces, non-living objects), the health status of the subjects, and the subjects' head pose. We study how these differences affect the performance of V-TEMP by conducting extensive tests to evaluate the effect of these differences. We also estimate the mean subject head motion for each video in a dataset by calculating the head pose for each subject using a 3D face tracker

(ZFace), to analyze the performance of V-TEMP with videos having different levels of subject head motion [78], [79], [80]. Table II shows a summary of the datasets utilized in our study.

To test whether we can indeed identify a biological temperature signal, we utilize a subset of videos of non-living objects from the Objectron dataset [81] and analyze the performance of our approach. The Objectron dataset consists of object-centric, pose-annotated video clips, where the camera moves around the object, capturing a video from different angles. We randomly sampled 60 videos recorded in diverse environments of non-living objects belonging to six classes (book, bottle, laptop, cerealbox, camera and cup) to analyze the efficacy of inferring a biological temperature signal from a video. Hypothesizing that non-living objects do not show a statistically significant difference in the angular reflectance distribution as compared to living skin (from four different datasets), we investigate whether we can infer a *HOT factor* (which implies the detection of a biological/temperature signal) for non-living videos from the sampled Objectron dataset. The mean camera rotation of videos from the sampled Objectron dataset is found to be $35.1°$. We utilize videos from four datasets consisting of subjects seated in a controlled laboratory setting to obtain the angular reflectance distribution of real human skin to investigate the determination of the *HOT factor* as compared to non-living objects (from the Objectron dataset). We utilized videos from three existing facial video datasets, the BP4D+ [82], the PSU-HR [83] and the 3DFAW [68] and one dataset we collected for this study. The BP4D+ dataset consists of 1400 facial videos of 140 subjects performing 10 tasks demonstrating a variety of emotions such as anger, fear and surprise. The subjects are seated indoors in a laboratory setting under controlled lighting settings with minimal environmental variations. The mean subject head motion for the BP4D+ dataset was calculated to be $2.7°$, showing that there is very small head motion in the dataset. The BP4D+ dataset consists of videos of healthy subjects with non-elevated skin temperature. We randomly sampled 32 videos from the BP4D+ dataset to ensure consistency with the other datasets. The PSU-HR dataset consists of 24 subjects seated indoors with minimal head angular deviation. The videos are collected with the approval of and in accordance with the relevant guidelines and regulations of the Pennsylvania State University's *Institutional Review Board* (IRB) [83]. Written informed consent was obtained from the subjects to record their facial video. The mean subject head motion was calculated to be $1.8°$. The PSU-HR
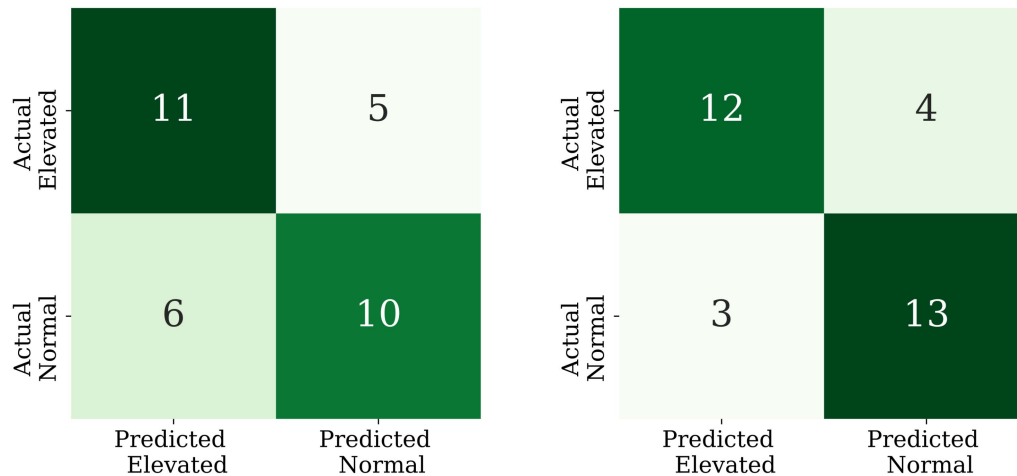
Fig. 7. Experimental results for V-TEMP in outside-the-lab conditions. Confusion matrices for videos sourced from the NEST-Wild dataset with lesser than 5° subject head angular deviation (left) and greater than 25° subject head angular deviation (right).

dataset consists of videos (recorded using the Surface 4 tablet) of healthy subjects with normal skin temperature, as measured using the Smarttemp FDA-approved Bluetooth temperature monitoring system. The 3DFAW dataset consists of 26 subjects performing the full range of left-to-right (yaw angular) head motion in an indoor, controlled setting with minimal lighting and environmental variations. The mean subject head motion is found to be 22.7°, which is significantly higher compared to the BP4D+ and the PSU-HR datasets. The 3DFAW dataset consists of videos of healthy subjects with normal skin temperature. Due to the lack of videos of subjects with elevated skin temperature in the BP4D+, the 3DFAW and the PSU-HR datasets, we developed an experimental protocol to record facial videos of subjects with non-elevated and artificially elevated skin temperature in a controlled environment (Section: Experimental Protocol to Develop the NEST-Lab Dataset) to create the Normal and Elevated Skin Temperature - Laboratory (NEST-Lab) dataset. The NEST-Lab dataset consists of 32 videos of subjects with normal skin temperature and 32 videos of the same subjects with elevated skin temperature caused by artificial skin heating as described in Section: Experimental Protocol to Develop the NEST-Lab Dataset. Performing a quality analysis, we excluded the videos of 4 subjects due to insufficient video length, providing us with 28 videos of subjects with normal skin temperature and 28 videos of the same subjects with elevated skin temperature. The videos in the NEST-Lab dataset are recorded in temperature-controlled laboratory settings (74° F) which minimizes the effect of environmental factors such as sunlight and wind speed on the skin temperature. The ambient lighting conditions and the background (white background) are maintained constant. The mean subject head motion for the NEST-Lab dataset is 34.3°. The ground truth temperature is recorded using the FLIR One Pro thermal camera as described in Section: Experimental Protocol to Develop the NEST-Lab Dataset. The skin tone distribution of subjects in the NEST-Lab dataset shows the similarity in the mean pixel intensities of subjects with non-elevated skin temperature and subjects with elevated skin temperature (Fig. 6). Fig. 7 shows the distribution of skin tone in subjects from the

NEST-Lab dataset classified into ten skin tone categories as defined by the Monk scale [84]. The study was conducted with the approval of and in accordance with the relevant guidelines and regulations of The Carnegie Mellon University's *Institutional Review Board* (IRB). Written informed consent was obtained from the subjects to record their facial videos. We analyze the performance of V-TEMP in controlled laboratory conditions using videos from the NEST-Lab dataset consisting of subjects with both elevated and non-elevated skin temperature.

Having analyzed the performance of V-TEMP in controlled settings involving real humans with both elevated and non-elevated skin temperature, we explore the performance in outside-the-lab environments. We curated 56 videos from online sources to create the Normal and Elevated Skin Temperature - Wild (NEST-Wild) dataset. The NEST-Wild dataset consists of 28 subjects with elevated skin temperature (15 subjects indicated that they were recovering from COVID-19) and 28 subjects with normal skin temperature. The NEST-Wild dataset consists of subjects recording themselves in an indoor setting, which minimizes the effect of environmental factors such as sunlight and wind speed on the skin temperature. The ambient lighting conditions in the NEST-Wild dataset are not controlled (unlike the NEST-Lab dataset) and are representative of outside-the-lab conditions. We focused on curating videos where the subjects demonstrated a high range of head motion and reported their temperature using a thermometer. The mean subject head motion for the NEST-Wild dataset is 14.9°. We consider the self-reported temperature of the subjects (using a thermometer) as the ground truth for classifying the subjects as having normal or elevated skin temperature (Supplementary Information). The study was conducted with the approval of and in accordance with the relevant guidelines and regulations of The Carnegie Mellon University's *Institutional Review Board* (IRB). The videos were curated according to the Fair Use Policy of the video sharing platform [90].

We ensure that all the videos in the datasets contain faces of a single subject (or single object in the frame for the Objectron dataset) with no occlusion. We define a video recorded in an

indoor setting, with a single person in the frame exhibiting a range of left-to-right head movement as an effective video to apply V-TEMP. Fig. 11 outlines the steps to generate a video suitable for analysis with V-TEMP. The face of the subject in the input video is tracked using a face tracker that can extract dense facial landmarks and also provide head pose information. The video is further processed to remove frames where the subject leaves the video frame, causing a loss of face tracking.

Having established that reflectance properties of living skin are necessary for V-TEMP, we explore whether non-skin regions of the face can also aid in the detection of a biological temperature signal. We hypothesize that there is no statistically significant difference between the reflectance distributions of skin and non-skin facial regions. To this end, we perform a hypothesis test to analyze the reflectance distributions of facial skin regions (nose) and non-skin regions (eyes) for videos from the BP4D+, the 3DFAW, the PSU-HR and the NEST-Lab datasets. The null and alternate hypotheses are constructed for each dataset as.

$$H_0 : R_{s_k} = R_{e_k} \tag{24}$$

$$H_a : R_{s_k} \neq R_{e_k} \tag{25}$$

where R_{s_k} and R_{e_k} are the distributions of the reflectance for facial skin regions (nose) and non-skin regions (eyes) respectively and $k \in$ [BP4D+, 3DFAW, PSU-HR, NEST-Lab]. Using power analysis, we determine a significance level of $\alpha = 0.1$. Performing the two sample Kolmogorov-Smirnov test, we obtain the p-values (for both $\alpha = 0.1$ and $\alpha = 0.05$) shown in Table IV, rejecting the null hypothesis and demonstrating that there is a statistically significant difference in the angular reflectance distributions of facial skin regions and non-skin regions. We observed that we obtain a *HOT factor* for all the videos while using the facial skin region, while we do not obtain the *HOT factor* for the videos while using the non-skin facial region. This shows that the efficacy in detecting a biological temperature signal depends only on facial skin-like reflectance properties.

Having tested the determination of the *HOT factor* from real human facial skin-like regions for videos of subjects with non-elevated skin temperature we explore the determination of the *HOT factor* for videos of subjects with elevated skin temperature in laboratory settings. We hypothesize that there is no statistically significant difference between the mean reflectance of videos of subjects with elevated skin temperature and videos of subjects with non-elevated skin temperature. To this end, we perform the hypothesis test as follows.

$$H_0 : \hat{R}_{elevated} = \hat{R}_{non-elevated} \tag{26}$$

$$H_a : \hat{R}_{elevated} \neq \hat{R}_{non-elevated} \tag{27}$$

where $\hat{R}_{elevated}$ is the distribution of mean reflectance of videos of subjects with elevated skin temperature and $\hat{R}_{non-elevated}$ is the distribution of mean reflectance of videos of subjects with non-elevated skin temperature from the NEST-Lab dataset. Using power analysis, we determine a significance level of

$\alpha = 0.1$. Performing the two sample Kolmogorov-Smirnov test, we obtain a p-value (for both $\alpha = 0.1$ and $\alpha = 0.05$) of 0.40, failing to reject the null hypothesis. This shows that there is no significant difference in the mean reflectance distributions of subjects with elevated skin temperature and subjects with non-elevated skin temperature. This result reveals that the raw mean skin reflectance value cannot differentiate between elevated and non-elevated skin and hence we leverage the angular dependence of the skin reflectance to detect elevated skin temperature. Further, we obtain a *HOT factor* for each video in the NEST-Lab dataset, both for subjects with non-elevated as well as elevated skin temperature. We now evaluate the performance of V-TEMP in classifying a subject as having elevated/non-elevated skin temperature using videos from the NEST-Lab dataset.

Power analysis was performed for each hypothesis test to determine the significance level α. Due to the non-parametric nature of the Kolmogorov-Smirnov test, we calculate the correlation between the two groups using the formula suggested by Field [69] and Rosenthal [70].

$$r = \frac{Z}{\sqrt{N}} \tag{28}$$

where r is the correlation between the two groups, Z is the KS test statistic and N is the number of samples in the group. Using results from Rosenthal [71], the correlation can be converted to Cohen's d [72] to determine the effect size. For the hypothesis test examining the difference between reflectance distributions of living and non-living objects (22), the weak correlation between the two distributions gives an effect size of 0.2. With a sample size ratio of 3, ($\approx$30 living samples (for each of BP4D+, 3DFAW, PSU-HR and non-elevated subjects from the NEST-Lab dataset) compared to 90 non-living samples), we obtain a significance level, α of 0.1, obtaining a statistical power of 0.8. For the hypothesis test examining the difference between reflectance distributions of facial skin regions and facial non-skin regions (24), the weak correlation between the two distributions gives an effect size of 0.2. With a sample size ratio of 1, ($\approx$30 living samples (for each of BP4D+, 3DFAW, PSU-HR and non-elevated subjects from the NEST-Lab dataset)), we obtain a significance level, α of 0.1, obtaining a statistical power of 0.75. For the hypothesis test examining the difference between the mean reflectance distributions of videos of subjects with elevated skin temperature as compared to videos of subjects with non-elevated skin temperature from the NEST-Lab dataset, (26) the strong correlation between the two distributions gives an effect size of 0.3. With a sample size ratio of 1 (28 videos), we obtain a significance level, α of 0.1, obtaining a statistical power of 0.75.

Having established that V-TEMP leverages reflectance properties of real facial skin to infer the presence of elevated skin temperature, we tested V-TEMP to determine optimal parameters such as face region of interest (obtained from the ZFace tracker), size of region of interest and threshold temperature using videos from the NEST-Lab dataset. Subject independent K-fold validation with $K = 5$ folds was performed to test the

TABLE III
EVALUATION OF V-TEMP WITHIN AND ACROSS DATASETS WITH ADDITIONAL SENSITIVITY ANALYSIS

Training Dataset ⇒ Testing Dataset	Type of study	Videos with non-elevated skin temperature	Videos with elevated skin temperature	F1 Score
NEST-Lab (0° - 45°) ⇒ NEST-Lab (0° - 45°)	Within dataset analysis	28	28	0.81
NEST-Lab (0° - 45°) ⇒ NEST-Wild (0° - 45°)	Across dataset analysis	28	28	0.76
NEST-Lab (0° - 45°) ⇒ NEST-Lab (> 25°)	Sensitivity analysis	16	16	0.85
NEST-Lab (0° - 45°) ⇒ NEST-Lab (< 5°)	Sensitivity analysis	16	16	0.65
NEST-Lab (0° - 45°) ⇒ NEST-Wild (> 25°)	Sensitivity analysis	16	16	0.78
NEST-Lab (0° - 45°) ⇒ NEST-Wild (< 5°)	Sensitivity analysis	16	16	0.67

efficacy of V-TEMP. In each iteration of K-fold cross validation, we employed 4 folds (3 folds for training and 1 fold for validation) for determining the optimal values for each parameter and tested on the held-out fold. The optimal facial regions for classification were identified to be the region between the eyes, as well as the nose bridge. The optimal size of the facial region of interest was found to be a square of side 11 pixels. We then report the F1 score obtained across 5 iterations of K-fold validation, by aggregating the number of true positives and true negatives obtained in each iteration. We also utilize the mode of the optimal parameters obtained during each iteration of K-fold validation to test V-TEMP on different datasets (Table III).

V-TEMP obtained an F1 score of 0.81 for the NEST-Lab dataset (Fig. 12) demonstrating the efficacy of V-TEMP in identifying subjects with elevated skin temperature in controlled laboratory settings. Additionally, to investigate the dependence of V-TEMP on subject head angular deviation, we extracted 32 video segments with head angular deviation less than 5° (small head angular deviation) and 32 video segments with head angular deviation greater than 25° (large head angular deviation) from the NEST-Lab dataset. As the subjects perform a continuous head motion from left to right, we can extract the required range of head angular deviation, i.e., 5° and 25°. We ensure an equal proportion of subjects with non-elevated and subjects with elevated skin temperature for both groups of subject head angular deviation. Using the optimal parameters obtained during K-fold validation on the original NEST-Lab dataset, we observe the F1 score of 0.85 for the video segments with larger head angular deviation to be significantly higher than the F1 score of 0.65 for the video segments with smaller head angular deviation (Table III). The corresponding confusion matrices for the video segments with large and small subject head angular deviation are shown in Fig. 13. The F1 score of 0.65 for video segments of subjects with small head angular deviation shows that the algorithm performs closer to a random binary classifier for video segments with small head angular deviation. The F1 score of 0.85 for video segments of subjects with large head angular deviation shows that the proposed approach demonstrates improved classification performance for video segments of subjects with larger head angular deviation. The F1 score of 0.85 obtained for video segments with higher angular deviation is different from the F1 score obtained for the full NEST-Lab dataset (F1 score of 0.81), as we are only evaluating on a subset of video segments showing higher angular deviation (> 25°). The highest

F1 score is observed with video segments with a subject head angular deviation larger than 25°.

To investigate the generalizability of V-TEMP, we evaluate the performance on the NEST-Wild dataset, without retraining the optimal parameters obtained during K-fold validation on the NEST-Lab dataset. We obtain an F1 score of 0.76 for the NEST-Wild dataset, using the optimal parameters obtained during K-fold validation for the NEST-Lab dataset. This result demonstrates the robustness of V-TEMP in detecting elevated skin temperature (Fig. 12). This result demonstrates the capability of V-TEMP in inferring a temperature from subjects in laboratory conditions as well as outside-the-lab conditions. We examined the performance of V-TEMP on video segments with head angular deviation less than 5° (small head angular deviation) and 32 video segments with head angular deviation greater than 25° (large head angular deviation) from the NEST-Wild dataset. We utilized the optimal parameters obtained during K-fold validation on the original NEST-Lab dataset to examine the performance of V-TEMP on video segments with larger head angular deviation (>25°) and smaller head angular deviation (<5°) from the NEST-Wild dataset. We observe the F1 score of 0.78 for the video segments with larger head angular deviation to be significantly higher than the F1 score 0.67 for the video segments with smaller head angular deviation (Table V) from the NEST-Wild dataset, without explicit retraining of the optimal parameters obtained for the NEST-Lab dataset. The corresponding confusion matrices for the video segments with large and small subject head angular deviation are shown in Fig. 7. This result demonstrates that V-TEMP performs better for videos with larger subject head angular deviation (>25°) in laboratory conditions as well as outside-lab conditions.

Rise in skin temperature can indicate the presence of an infection in the body. Due to the increased prevalence of diseases causing fever (such as COVID-19), detecting elevated skin temperature is critical. Such a system for detecting elevated skin temperature can be installed in areas with a high footfall for real-time mass screening of people. However, skin temperature can also be influenced by environmental factors such as sunlight and wind speed. A video taken indoors will eliminate the aforementioned environmental factors. The videos in the datasets chosen for this study are all recorded indoors, which minimizes the influence of environmental conditions. Skin temperature also depends on the physiology of the person such as

the age, gender and the activity level. Analyzing the effect of factors such as age, gender and activity level is out of the scope of this study due to the difficulty involved in obtaining such information from videos shared online.

IV. CONCLUSION

In this study, we present an algorithm (V-TEMP) to estimate the presence of elevated skin temperature using a facial video of a person, captured with a standard camera found in most mobile phones. Based on angular reflectance distributions, we find that subject head rotation can be employed to detect elevated skin temperature of the person and propose V-TEMP accordingly. Through statistical hypothesis testing, we discover that V-TEMP can uniquely detect a biological temperature signal from living facial skin. We find that V-TEMP can effectively classify healthy subjects and subjects with elevated skin temperature, even in environmental conditions typically found outside-the-lab. Interestingly, we find that the facial videos of subjects with smaller head angular deviation depreciate the applicability of the algorithm. The requirement of larger subject head motion is also beneficial for the deployment of our method in outside-the-lab settings, as videos recorded in outside-the-lab settings typically contain a large amount of head motion. Another key contribution of the study is the curation of facial video datasets consisting of subjects with elevated skin temperature in both laboratory conditions and outside-the-lab conditions (NEST-Lab and NEST-Wild). The development of the NEST-Lab and the NEST-Wild datasets is crucial due to the inadequacy of subjects with elevated skin temperature in existing video datasets. Our work lays a foundation for elevated skin temperature detection using an RGB camera. The use of an RGB video instead of an infrared video increases the deployability of our approach due to the ubiquity of devices that can process RGB videos, such as mobile phones and computers. The use of RGB videos is also instrumental in the development of temperature screening applications. This work provides motivation for the development of video-based, non-infrared temperature screening applications which are especially useful in remote areas lacking access to specialized resources such as infrared cameras. In the future, the detection of the absolute core body temperature with video-based approaches can be studied, which can aid the diagnosis of diseases.

Data Availability: The datasets collected in this study will be uploaded to https://github.com/AiPEX-Lab/skin-temp.

Code Availability: All custom code used for temperature analysis will be uploaded to https://github.com/AiPEX-Lab/skin-temp. The ZFace tracking library can be found at http://zface.org.

Competing Interests: The authors declare no competing interests.

ACKNOWLEDGMENT

REFERENCES

[1] N. Charkoudian, "Skin blood flow in adult human thermoregulation: How it works, when it does not, and why," *Mayo Clinic Proc.*, vol. 78, pp. 603–612, 2003.

[2] H. Daanen et al., "Covid-19 and thermoregulation-related problems: Practical recommendations," *Temperature*, vol. 8, no. 1, pp. 1–11, 2021.

[3] C. Or and V. Duffy, "Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement," *Occup. Ergonom.*, vol. 7, no. 2, pp. 83–94, 2007.

[4] A. Helmy, M. Holdmann, and M. Rizkalla, "Application of thermography for non-invasive diagnosis of thyroid gland disease," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1168–1175, Mar. 2008.

[5] I. Buss et al., "Aetiology of fever in returning travellers and migrants: A systematic review and meta-analysis," *J. Travel Med.*, vol. 27, no. 8, 2020, Art. no. taaa207.

[6] H. Hymczak et al., "Core temperature measurement—Principles of correct measurement, problems, and complications," *Int. J. Environ. Res. Public Health*, vol. 18, no. 20, 2021, Art. no. 10606.

[7] A. El-Radhi et al., *Clinical Manual of Fever in Children*. Berlin, Germany: Springer, 2009.

[8] J. Craig et al., "Temperature measured at the axilla compared with rectum in children and young people: Systematic review," *BMJ*, vol. 320, no. 7243, pp. 1174–1178, 2000.

[9] P. Valle et al., "A comparative study of infrared tympanic thermometry and rectal mercury thermometry," *Scand. J. Infect. Dis.*, vol. 31, no. 1, pp. 105–106, 1999.

[10] J. Craig et al., "Infrared ear thermometry compared with rectal thermometry in children: A systematic review," *Lancet*, vol. 360, no. 9333, pp. 603–609, 2002.

[11] G. Dell'Isola et al., "Noncontact body temperature measurement: Uncertainty evaluation and screening decision rule to prevent the spread of COVID-19," *Sensors*, vol. 21, no. 2, 2021, Art. no. 346.

[12] F. Rodriguez-Lozano et al., "Non-invasive forehead segmentation in thermographic imaging," *Sensors*, vol. 19 no. 19, 2019, Art. no. 4096.

[13] H. Chen et al., "Investigation of the impact of infrared sensors on core body temperature monitoring by comparing measurement sites," *Sensors*, vol. 20, no. 10, 2020, Art. no. 2885.

[14] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, 2007, Art. no. R1.

[15] J. Allen et al., "Microvascular blood flow and skin temperature changes in the fingers following a deep inspiratory gasp," *Physiol. Meas.*, vol. 23, no. 2, 2002, Art. no. 365.

[16] N. Ruben, "Remote heart rate estimation using consumer-grade cameras," *All Graduate Plan B and other Reports*, vol. 463, 2015.

[17] A. Sagaidachnyi et al., "Correlation of skin temperature and blood flow oscillations," *Proc. SPIE*, vol. 8337, 2012, Art. no. 83370A.

[18] A. Sagaidachnyi et al., "Determination of the amplitude and phase relationships between oscillations in skin temperature and photoplethysmography-measured blood flow in fingertips," *Physiol. Meas.*, vol. 35, no. 2, 2014, Art. no. 153.

[19] F. Sanchez-Marin et al., "Novel approach to assess the emissivity of the human skin," *J. Biomed. Opt.*, vol. 14 no. 2, 2009 Art. no. 024006.

[20] C. Liu et al., "Limitations of forehead infrared body temperature detection for fever screening for severe acute respiratory syndrome," *Infection Control Hosp. Epidemiol.*, vol. 25, no. 12, pp. 1109–1111, 2004.

[21] D. Ng et al., "A brief report on the normal range of forehead temperature as determined by noncontact, handheld, infrared thermometer," *Amer. J. Infection Control*, vol. 33, no. 4, pp. 227–229, 2005.

[22] C. Cheng et al., "Increasing mean skin temperature linearly reduces the core-temperature thresholds for vasoconstriction and shivering in humans," *J. Amer. Soc. Anesthesiologists*, vol. 82, no. 5, pp. 1160–1168, 1995.

[23] J. Avner, "Acute fever," *Pediatr. Rev.*, vol. 30, no. 1, p. 5, 2009.

[24] K. Şahin, "Thermoregulation: Normal physiology, anesthetic effects, factors disrupting thermoregulation in the perioperative period," *Health Sci.*, Art. no. 179.

[25] M. Masè et al., "Insight into the use of tympanic temperature during target temperature management in emergency and critical care: A scoping review," *J. Intensive Care*, vol. 9, no. 1, pp. 1–18, 2021.

[26] A. Pandharkar and G. Thote, "Approach to fever and its homoeopathic management," *J. Med. Pharmaceut. Innov.*, vol. 8, no. 40, pp. 4–10, 2021.

[27] V. Shusterman et al., "Spontaneous skin temperature oscillations in normal human subjects," *Amer. J. Physiol.-Regulatory, Integrative Comp. Physiol.*, vol. 273, no. 3, pp. R1173–R1181, 1997.

[28] J. Foster et al., "Non-contact infrared assessment of human body temperature: The journal temperature toolbox. Temperature," *Temperature*, vol. 8, pp. 306–319, 2021.

[29] S. Lim et al., "An unsupervised machine learning model for discovering latent infectious diseases using social media data," *J. Biomed. Inform.*, vol. 66, pp. 82–94, 2017.

[30] A. Michel et al., "In vivo measurement of mid-infrared light scattering from human skin," *Biomed. Opt. Exp.*, vol. 4, no. 4, pp. 520–530, 2013.

[31] Y. Gao and G. Tian, "Emissivity correction using spectrum correlation of infrared and visible images," *Sensors Actuators A: Phys.*, vol. 270, pp. 8–17, 2018.

[32] A. Krishnaswamy and G. Baranoski, "A biophysically-based spectral model of light interaction with human skin," *Comput. Graph. Forum*, vol. 23, pp. 331–340, 2004.

[33] A. Sohaib et al., "BRDF of human skin in the visible spectrum," *Sensor Rev.*, vol. 37, pp. 390–395, 2017.

[34] T. Chen et al., "Hyperspectral modeling of skin appearance," *ACM Trans. Graph.*, vol. 34, no. 3, pp. 1–14, 2015.

[35] J. Gole and D. Dixon, "Suggested correlation between the visible photoluminescence and the Fourier transform infrared spectrum of a porous silicon surface," *J. Phys. Chem. B*, vol. 101, no. 41, pp. 8098–8102, 1997.

[36] C. Gatebe et al., "Sensitivity of off-nadir zenith angles to correlation between visible and near-infrared reflectance for use in remote sensing of aerosol over land," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 805–819, Apr. 2001.

[37] I. Meglinski and S. Matcher, "Computer simulation of the skin reflectance spectra. Computer methods and programs in biomedicine," *Comput. Methods Prog. Biomed.* vol. 70, no. 2, pp. 179–186, 2003.

[38] P. Forsström et al., "Multi-angular reflectance spectra of small single trees," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112302.

[39] P. Brimmer et al., "Angular dependence of diffuse reflectance infrared spectra. Part I: FT-IR spectrogoniophotometer," *Appl. Spectrosc.*, vol. 40, no. 2, pp. 258–265, 1986.

[40] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 463–477, Mar. 2016.

[41] J. West and A. Luks, *West's Pulmonary Pathophysiology: The Essentials*. Philadelphia, PA, USA: Lippincott Williams Wilkins, 2021.

[42] L. Fleisher et al., "Thermoregulation and heart rate variability," *Clin. Sci.*, vol. 90, no. 2, pp. 97–103, 1996.

[43] I. Jeong et al., "Effects of skin surface temperature on photoplethysmograph," *J. Healthcare Eng.*, vol. 5, no. 4, pp. 429–438, 2014.

[44] S. Khan et al., "Comparative accuracy testing of non-contact infrared thermometers and temporal artery thermometers in an adult hospital setting," *Amer. J. Infection Control*, vol. 49, no. 5, pp. 597–602, 2021.

[45] V. Betta et al., "An assessment of infrared tympanic thermometers for body temperature measurement," *Physiol. Meas.*, vol. 18, no. 3, 1997, Art. no. 215.

[46] E. Y. K. Ng, W. Muljo, and B. S. Wong, "Study of facial skin and aural temperature," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 3, pp. 68–74, May/Jun. 2006.

[47] I. Pušnik et al., "IR ear thermometers: What do they measure and how do they comply with the EU technical regulation?," *Physiol. Meas.*, vol. 25, no. 3, 2004, Art. no. 699.

[48] A. Shajkofci, "Correction of human forehead temperature variations measured by non-contact infrared thermometer," *IEEE Sensors J.*, vol. 22, no. 17, pp. 16750–16755, Sep. 2022.

[49] M. Cho et al., "Comparing axillary and rectal temperature measurements in very preterm infants: A prospective observational study," *Neonatol.*, vol. 118, no. 2, pp. 180–186, 2021.

[50] H. Kocoglu et al., "Infrared tympanic thermometer can accurately measure the body temperature in children in an emergency room setting," *Int. J. Pediatr. Otorhinolaryngol.*, vol. 65, no. 1, pp. 39–43, 2002.

[51] D. Ng et al., "Non-contact infrared thermometry temperature measurement for screening fever in children," *Ann. Trop. Paediatr.*, vol. 25, no. 4, pp. 267–275, 2005.

[52] J. Chaglla et al., "Measurement of core body temperature using graphene-inked infrared thermopile sensor," *Sensors*, vol. 18, no. 10, 2018, Art. no. 3315.

[53] A. Sharma and A. Yadav, "Image processing based body temperature estimation using thermal video sequence," in *Proc. Int. Conf. Comput. Methodol. Commun.*, 2017, pp. 846–852.

[54] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. I–I.

[55] J.W. Chin et al., "An infrared thermography model enabling remote body temperature screening up to 10 meters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021, pp. 3870–3876.

[56] A. Revanur et al., "The first vision for vitals (v4v) challenge for non-contact video-based physiological estimation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2760–2767.

[57] A. Revanur et al., "Instantaneous physiological estimation using video transformers," 2022, *arXiv:2202.12368*.

[58] Edward P. Tryon, *Encyclopedia of Physical Science and Technology*, 3rd ed. Cambridge, MA, USA: Academic Press, 2003.

[59] J. Stefan, "On the relation between heat radiation and temperature," *Ann. Phys. Annalen der Physik und Chemie*, vol. 258, no. 6, pp. 291–294, 1884.

[60] L. Boltzmann, "Erivation of Stefan's law, concerning the dependency of heat radiation on temperature, from the electromagnetic theory of light Proceedings of the Imperial Philosophical Academy [of Vienna]: Mathematical and Scientific Class," vol. 79, pp. 391–428, 1879.

[61] I. Fernández-Cuevas et al., "Classification of factors influencing the use of infrared thermography in humans: A review," *Infrared Phys. Technol.*, vol. 71, pp. 28–55, 2015.

[62] R. Usamentiaga et al., "Infrared thermography for temperature measurement and non-destructive testing," *Sensors*, vol. 14, no. 7, pp. 12305–12348, 2014.

[63] L. MacDonald, "Colour and directionality in surface reflectance," in *Proc. Conf. Artif. Intell. Simul. Behav.*, 2014.

[64] Y. Sato and K. Ikeuchi, "Reflectance analysis under solar illumination," in *Proc. Workshop Phys.-Based Model. Comput. Vis.*, 1995, pp. 180–187.

[65] T. Cheng et al., "Curvature effect quantificaiton for in-vivo IR thermography," in *Proc. ASME Int. Mech. Eng. Congr. Expo.*, 2012, pp. 127–133.

[66] Y. Arieli et al., "Correction of the angular emissivity of human skin for clinical thermal imaging," *Imag. Med.*, vol. 9, no. 4, pp. 103–108, 2017.

[67] X. Zhang et al., "A high-resolution spontaneous 3D dynamic facial expression database," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2013, pp. 1–6.

[68] L. Jeni et al., "3D dense face reconstruction from video (3DFAW-video) challenge," in *Proc. 2nd Workshop Challenge 3D Face Alignment Wild Dense Reconstruction Video, Conjunction IEEE Int. Conf. Comput. Vis.*, 2019.

[69] A. Field, *Reliability Analysis. Discovering Statistics Using SPSS*, 2nd ed.Thousand Oaks, CA, USA:Sage Publishing, 2005.

[70] R. Rosenthal, "Meta-analytic procedures for social research," *Appl. Social Res. Methods Ser.*, 1991.

[71] R. Rosenthal et al., "Parametric measures of effect size," in *The Handbook of Res. Synth.*, vol. 621. 1994, pp. 231–244.

[72] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Cambridge, MA, USA: Academic Press, 2013.

[73] A. Dasari et al., "Evaluation of biases in remote photoplethysmography methods," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–13, 2021.

[74] H. Apa et al., "Clinical accuracy of tympanic thermometer and noncontact infrared skin thermometer in pediatric practice: An alternative for axillary digital thermometer," *Pediatr. Emerg. Care*, vol. 29, no. 9, pp. 992–997, 2013.

[75] M. Tay et al., "Comparison of infrared thermal detection systems for mass fever screening in a tropical healthcare setting," *Public Health*, vol. 129, no. 11, pp. 1471–1478, 2015.

[76] S. Hsiao et al., "Measurement of body temperature to prevent pandemic COVID-19 in hospitals in Taiwan: Repeated measurement is necessary," *J. Hosp. Infection*, vol. 105, pp. 360–361, 2020.

[77] K. Khaksari et al., "Review of the efficacy of infrared thermography for screening infectious diseases with applications to covid-19," *J. Med. Imag.*, vol. 8 no. S1, 2021 Art. no. 010901.

[78] I. O. Ertugrul et al., "AFAR: A deep learning based tool for automated facial affect recognition," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2019, pp. 1–1.

[79] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3D face alignment from 2D videos in real-time," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2015, pp. 1–8.

[80] L. Jeni et al., "Dense 3D face alignment from 2D video for real-time use," *Image Vis. Comput.*, vol. 58, pp. 13–24, 2017.

[81] A. Ahmadyan et al., "Objectron: A large scale dataset of object-centric videos in the wild with pose annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 7822–7831.

[82] Z. Zhang et al., "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3438–3446.

[83] S. Prakash and C. Tucker, "Bounded Kalman filter method for motion-robust, non-contact heart rate estimation," *Biomed. Opt. Exp.*, vol. 9, no. 2, pp. 873–897, 2018.

[84] E. Monk, "Monk skin tone scale,".

[85] F. Bartell et al., "The theory and measurement of bidirectional reflectance distribution function (BRDF) and bidirectional transmittance distribution function (BTDF)," *Radiat. Scattering Opt. Syst.*, vol. 257, pp. 154–160, 1981.

[86] S. Marschner et al., "Image-based bidirectional reflectance distribution function measurement," *Appl. Opt.*, vol. 39, no. 16, pp. 2592–2600, 2000.

[87] J. Kim et al., "Temporal changes in skin temperature caused by wet and dry heat in healthy volunteers," *J. Phys. Ther. Sci.*, vol. 24, no. 12, pp. 1209–1213, 2012.

[88] M. Lopes et al., "In vivo confocal raman spectroscopic analysis of the effects of infrared radiation in the human skin dermis," *Photochem. Photobiol.*, vol. 93, no. 2, pp. 613–618, 2017.

[89] "Infrared industrial heat incandescent," 2022.

[90] "YouTube Fair Use Policy." [Online]. Available: https://www.youtube.com/howyoutubeworks/policies/copyright/#fair-use