"Reading Between the Heat": Co-Teaching Body Thermal Signatures for Non-intrusive Stress Detection

YI XIAO, Syracuse University, USA
HARSHIT SHARMA, Syracuse University, USA
ZHONGYANG ZHANG, University of California San Diego, USA
DESSA BERGEN-CICO, Syracuse University, USA
TAUHIDUR RAHMAN, University of California San Diego, USA
ASIF SALEKIN, Syracuse University, USA

Stress impacts our physical and mental health as well as our social life. A passive and contactless indoor stress monitoring system can unlock numerous important applications such as workplace productivity assessment, smart homes, and personalized mental health monitoring. While the thermal signatures from a user's body captured by a thermal camera can provide important information about the "fight-flight" response of the sympathetic and parasympathetic nervous system, relying solely on thermal imaging for training a stress prediction model often lead to overfitting and consequently a suboptimal performance. This paper addresses this challenge by introducing ThermaStrain, a novel co-teaching framework that achieves high-stress prediction performance by transferring knowledge from the wearable modality to the contactless thermal modality. During training, ThermaStrain incorporates a wearable electrodermal activity (EDA) sensor to generate stress-indicative representations from thermal videos, emulating stress-indicative representations from a wearable EDA sensor. During testing, only thermal sensing is used, and stress-indicative patterns from thermal data and emulated EDA representations are extracted to improve stress assessment. The study collected a comprehensive dataset with thermal video and EDA data under various stress conditions and distances. ThermaStrain achieves an F1 score of 0.8293 in binary stress classification, outperforming the thermal-only baseline approach by over 9%. Extensive evaluations highlight ThermaStrain's effectiveness in recognizing stress-indicative attributes, its adaptability across distances and stress scenarios, real-time executability on edge platforms, its applicability to multi-individual sensing, ability to function on limited visibility and unfamiliar conditions, and the advantages of its co-teaching approach. These evaluations validate ThermaStrain's fidelity and its potential for enhancing stress assessment.

1 INTRODUCTION

Stress is an intense emotional phenomenon that can be triggered by external stressors or stimuli [36, 53, 71]. It elicits spontaneous physiological responses governed by our autonomic nervous system's "fight or flight" response [12]. These responses may include changes in skin temperature [26, 106], skin conductance [76], and other indicators. Chronic stress poses significant risks to both physical and mental health, emphasizing the importance of monitoring and managing stress [89, 94].

Smart wearables have been explored for stress sensing [9, 79, 84]. However, such wearable-based solutions typically require close proximity to the user's body or skin for capturing different physiological parameters (e.g., electrodermal activities - EDA, skin temperature), which can be burdensome and invasive for the users. Furthermore, their sensing scope is limited to the wearer, restricting their applicability in indoor environments with multiple occupants. Passive and contactless indoor stress monitoring, on the other hand, offers the potential to unlock numerous applications that are challenging to achieve through EDA or other wearable-based solutions. For instance, passive and contactless stress monitoring for elderly dementia patients [28, 59, 98] or employees in smart workplaces [5, 64, 66, 74] can provide valuable insights. These approaches can facilitate feedback on stress, including bio-feedback [117], interventions for stress management, enhancements in well-being, and customization of user experiences based on stress-related data [65]. Further discussion on specific application scenarios for contactless passive stress sensing is provided in Section 8.1.

To address the challenge, several RGB camera-based stress sensing systems [30, 32, 118] have been developed; however, their efficacy depends on lighting conditions and is fraught with privacy concerns [35, 83]. Similarly,

remote photoplethysmography (PPG) based on RGB cameras fails to work well under varying lighting conditions [19]. Balancing accuracy, privacy, and adaptability to environmental variations remains a significant challenge for developing stress sensing systems.

Infrared thermography, which utilizes thermal cameras for stress sensing, can offer a viable solution. Thermal cameras can capture changes in skin temperature [47], heart rate through facial skin blood flow [56], which are indicative of physiological stress responses [1, 90]. Unlike RGB cameras, thermal cameras are robust to different light conditions [3]. Prior works have shown promising human sensing results using thermal imaging in poor lighting and even at night [27, 55, 83]. Additionally, thermal videos/imaging are typically considered more privacy-preserving compared to RGB imaging, preventing the inadvertent exposure of environmental/contextual information like personal items, addresses, displayed documents, and content within photo frames, among others [18, 35, 83]. These attributes enhance the appeal of thermal cameras for stress sensing.

Several studies [3, 23, 87] have explored the use of thermal sensing to assess stress. However, the efficacy of stress detection achieved solely through thermal sensing is lower compared to leveraging other single modalities such as EEG [107], ECG [51], and PPG [41]. Recognizing this limitation, recent studies [20, 31, 119] are focusing on multi-modality approaches to increase the efficacy of thermal stress sensing. These approaches typically require combining thermal data with other physiological signals from different modalities, such as EEG, ECG, or PPG, which increases computational cost, user burden and thus limits scalability.

This paper explores *"whether a system that utilizes stress-indicative physiological signals (from wearable sensors) during model development but relies solely on thermal sensing during evaluation or deployment can outperform uni-modal thermal camera-based stress sensing approaches."* - Uni-modal approaches use only thermal information in all model development and evaluation phases, discussed in Section 3.2.1.

To address the above-mentioned question, we introduce *ThermaStrain*, a first-of-its-kind end-to-end co-teaching framework that enhances the efficacy of infrared thermography-based stress sensing. By incorporating electrodermal activity (EDA) sensing (collected from wearable) in the model training phase, which is a reliable method for measuring human stress response in real-time [4, 49], *ThermaStrain* improves the accuracy of stress detection. During training, the model utilizes EDA sensing to generate a stress-indicative latent representation from thermal videos, emulating the stress-indicative signal patterns obtained from a real wearable EDA sensor. During test/evaluation time, when only thermal sensing is available, the stress-indicative information extracted from the thermal videos using the emulated EDA representation is used for stress detection. By integrating EDA sensing and learning to extract EDA-guided stress-indicative information from thermal videos, *ThermaStrain* offers an accurate and non-intrusive solution. It is a pioneering approach that addresses the research question while maintaining the simplicity and effectiveness of thermal sensing in stress assessment.

The main contributions of this work are:

- The paper presents *ThermaStrain*, a novel co-teaching-based solution that surpasses existing uni-modal and co-teaching baselines in thermal stress sensing (Sections 6.3 and 6.5). What sets *ThermaStrain* apart is its ability to achieve superior performance (Section 6.8) using only thermal sensing during evaluation/testing, thus maintaining the non-intrusive appeal of thermal sensing. *ThermaStrain* solution opens up new possibilities for enhancing ubiquitous computing applications where non-intrusive stress assessment plays a crucial role in promoting well-being and optimizing experiences, such as workplace productivity assessment, smart home systems, and personalized and passive mental health monitoring, including depression [38].
- To our knowledge, no existing public/available dataset contains variable distance, full or partial body thermal camera data, and physiological parameters in different stress conditions. To overcome this limitation, we collected a comprehensive dataset consisting of infrared thermography sensing (thermal video data) and electrodermal activity (EDA) physiological parameter sensing data. The dataset was

gathered from 32 individuals who performed four distinct stress-inducing tasks, each associated with different stressors. Importantly, data was collected from varying distances of 5-11 feet. This dataset's unique characteristics, including the diverse set of stressors and variable distances, allow the paper to develop and evaluate models that are capable of generalizing to different distances (Section 6.7)) and stress situations (Section 6.4). De-identified data will be made public. By addressing the gap in available datasets, this work enables more robust and applicable research in thermal stress sensing.

- The paper presents thorough evaluations and discussions (Sections 7 and 8) that delve into the benefits of co-teaching (i.e., *ThermaStrain*'s approach) in developing an improved stress-sensing solution. Section 7's evaluations encompassed various aspects, including understanding how co-teaching facilitates better solution development and the effectiveness of *ThermaStrain* in extracting stress-indicative information from thermal frames. Furthermore, Section 8 delves deeply into the potential applications and challenges of deploying *ThermaStrain* in real-world scenarios. This encompasses scenarios with multiple individuals, limited visibility, and conditions that are unseen during training, such as camera angles, distances, postures, stress conditions, backgrounds, and ethical concerns. The evaluations establish the fidelity of *ThermaStrain* to the co-teaching paradigm and validate its ability to enhance stress sensing.

2 MOTIVATION AND VISION MODEL OF CO-TEACHING-BASED THERMAL STRESS SENSING

This section discusses the vision model of thermal sensing and the motivation or justification of the presented co-teaching solution.

Vision Model of the Infrared Thermography Sensing. While tracking the thermal signatures of a target object, the thermal energy reaching the thermal camera sensors is formulated by Kylili et al. [63] as:

$$I = I_{EM} + I_{REF} + I_{ATM} \tag{1}$$

Here I_{EM} is the energy emitted by the object, I_{REF} is the energy reflected by the surrounding and intercepted by the object, and I_{ATM} is a term that accounts for atmospheric influence due to attenuation of thermal radiation. Here, the camera determines the I_{REF} and I_{ATM} during calibration. This allows the sensor to get information about the amount of thermal energy emitted by the target object, which is human body in this paper's scope.

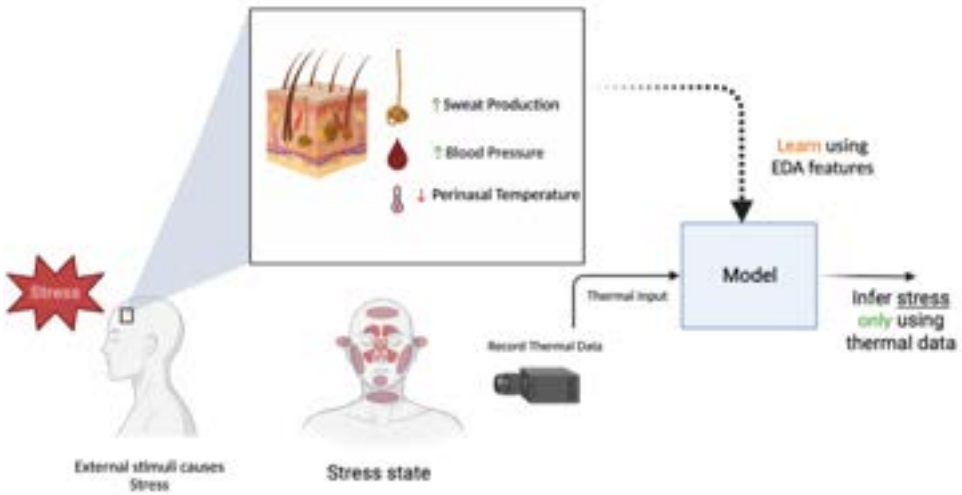


Fig. 1. Vision Model of Co-teaching based Thermal Sensing

Thermal-based Stress Sensing. While facing a stressful situation, our autonomic nervous system (ANS) changes the blood perfusion to the skin surface, which changes the skin temperature [26, 106]. Prior works [86, 106] have demonstrated that human body thermal signatures, i.e., skin temperature, particularly those from the face and neck regions [106], can provide an insight into human physiology under stressful conditions, that thermal camera captures through I_{EM}.

EDA-based Stress Sensing. Prior studies in psychophysiology have shown that electrodermal activity (EDA) or skin conductance is a gold standard to measure human stress response in real-time [4, 49]. While experiencing a stressful situation, there is an increase in the skin conductance levels, which is caused by the activation of the eccrine sweat glands [49, 61].

Limitation: Latency in Thermal Sensing. Thermal sensing can help to model the understanding of the stress response but cannot outperform the EDA-based assessment since skin conductance provides a rapid response profile (having a delay of 1-3 seconds from the stimulus onset) [10]. In contrast, thermal responses have a relatively higher latency of $4-5$ seconds from the stimulus onset [80, 106].

Co-teaching Goal. The Co-teaching goal is to learn the rapid patterns emerging through skin conductance, i.e., activation of eccrine sweat glands through the thermal energy measurement from the human body I_{EM}. The eccrine sweat glands are composed of a single tubular structure, and the volume of liquid in the tubular part of the eccrine glands increases when activated [44]. Studies [81, 92] have shown that an increase in sweating, i.e., water on the skin surface, results in the perception of lower temperature by the infrared thermography sensing (i.e., thermal camera information) than the thermal contact sensor (or actual skin temperature) [81, 92]. *Meaning there exists a latent thermal signature of skin conductance.* The co-teaching goal of the *ThermaStrain* approach is to teach thermal sensing modality to extract such latent patterns with the guidance of EDA modality, resulting in better stress assessment performance. Our presented end-to-end co-teaching approach effectively teaches such patterns, resulting in higher stress assessment performance from thermal sensing alone during testing or evaluation.

3 RELATED WORKS

Thermal imaging has shown promising results for physiology-based affective state and stress detection techniques in recent years [3, 20, 23, 87, 96, 119]. *ThermaStrain* solution builds upon two components of prior work (1) Thermal imaging for understanding electrodermal activity (EDA) responses. (2) Stress detection using thermal responses. These are discussed below:

3.1 Thermal Imaging for Measuring EDA Response

Recent works [19, 61, 86] showed that thermal imaging of skin areas with a high density of sweat glands like the palm or the perinasal regions [19] could be used to monitor the EDA response or the activation of the sweat glands. The activation of the sweat glands leads to a change in the skin temperature [3, 19], which is captured using the thermal camera. A recent study [86] found that there were high correlations between the galvanic skin response (GSR) extracted from the electrodermal activity (EDA) and the thermal signals extracted from the finger and perinasal region (correlation coefficient r=0.94 and r=0.96 respectively). Another work [61] studied the active pores on the skin surface using high-resolution thermal imaging and found a high correlation (correlation coefficient r=0.7) between the pore activation index measured using the thermal images [61] and skin conductance response measured from the finger. These works show the potential of thermal imaging for studying human physiological processes. Our work leverages the correlations between thermal responses and skin conductance, i.e., EDA, which these prior works have established.

3.2 Thermal Imaging for Stress Detection

Studies have utilized different physiological modalities like PPG [20, 41], ECG [51, 119], RGB image or video data [112, 119], EDA [102, 123], and thermal imaging [20, 23, 87, 119] to detect human stress. Recently, researchers have successfully used a combination of thermal imaging and different physiological sensors to detect individuals' affective state [3, 96], cognitive load [1], stress [20, 23, 87, 119] and even deception [85].

This section focuses on the prior literature on human stress detection, with a particular focus on thermal imaging based studies. These works can be divided into three strands based on their methodology.

3.2.1 Uni-Modal Approaches to Stress Detection: Uni-modal approaches in the literature use a single physiological modality like EEG [107], ECG [51], PPG [41] to detect human stress. In uni-modal thermal stress sensing scope, works like Cross et al. [23] used only thermal imaging to track regions of interest in the facial area to detect human stress using an LDA classifier, achieving 89.3% accuracy. Another work [87] used a thermal imaging-based approach to extract thermal maps corresponding to the facial, neck, and shoulder region for detecting positive and negative affective states with 90% accuracy by using statistical descriptors like average, minimum, maximum, and standard deviation and the difference between the minimum and the maximum temperature for the face, neck, and shoulder region in the thermal image frames. However, these approaches require the thermal cameras to be up close to an individual's face, hence have limited practical use in non-intrusive stress monitoring.

3.2.2 Multi-Modal Approaches to Stress Detection: Multi-modal approaches use two or more modalities for the human stress detection task. These techniques have also been widely studied in the literature.

Cho et al. [20] proposed a human stress measurement system using smartphone camera–based PPG and thermal video, achieving an average classification accuracy of 78.3% which outperformed the single modality baselines. Walambe et al. [112] explored early fusion and late fusion techniques to predict stress from posture, physiological, and video data. Their evaluation showed that early fusion outperformed late fusion by 5%. Can et al. [14] tried different schemes for modality fusion and found that using multiple modalities improved the performance of their stress detection systems in all scenarios. Ghosh et al. [31] converted data into Gramian Angular Field (GAF) before fusing multi-modality data, which can represent temporal correlations between each timestamp. They achieved significantly better performance than using raw data. Zhang et al. [119] fused ECG, voice, and RGB facial video for acute stress detection. Their ablation study showed that the overall performance was improved using ResNet 50 and Inflated 3D-CNN.

Finally, multi-modal machine learning approaches leverage information from multiple modalities and often achieve higher accuracy than uni-modal approaches. However, not all modalities may be available in real-life scenarios and can be costly and comparatively more invasive, limiting their practical use [120].

3.2.3 Co-teaching Approaches. To address the limitation of multi-modality, many studies have attempted to reconstruct missing modalities from existing ones to address the issue of missing modalities at inference time. These approaches fall under the domain of Co-Teaching. E.g., Zheng et al. [120] trained a prototype network to learn meta-sensory representations by modeling knowledge retention mechanisms. Rajan et al. [93] proposed a modality translator to translate the weak modality of strong modality, so that weak modality alone can achieve better performance during evaluation. Fortin et al. [29] proposed a multi-task learning framework that prepares multiple classifiers depending on the availability of modalities. Wang et al. [113] designed a Generative Adversarial Network (GAN) to reconstruct the missing modality. Li et al. [68] trained a Visual Hallucination Transformer that maps text to images and showed that visualizing scenes from the text can improve machine translation systems. This paper considers the multi-task learning [29] and 'Hallucination Transformer' [68] as baselines.

None of the above-discussed studies are on thermal imaging or stress. The closest state-of-the-art work to co-teaching on thermal imaging is StressNet [62], which obtained ECG attributes from thermal input and utilized the extracted ECG-relevant thermal embedding to predict stress. The study utilizes only closed facial thermal

frames, extracts ECG-relevant embedding using a ResNet, and captures temporal dynamics through an LSTM backbone. Even though it is not exactly co-teaching, due to its similarity to the concept, it is considered one of the baselines of this paper.

4 DESCRIPTION OF THE DATASET AND OUR DATA COLLECTION PROCEDURE

To address the lack of an available dataset containing variable distance, full or partial body thermal camera data, and physiological parameters in different stress conditions, the paper collected data in an indoor setting. Participants were engaged in various non-stress and stress-inducing tasks. The tasks were carefully designed in collaboration with a behavioral psychologist and approved by the X University Institutional Review Board (IRB) to ensure ethical compliance.

Participants: The participants in the dataset were 32 undergrad and graduate students enrolled at X University. They comprised 12 male and 20 female participants ($22 - 32$ years of age). All data were collected in a single laboratory visit, and participants signed informed consent before initiating the study. This limited age group may not be generalizable to older adults or children.

4.1 Sensing Modalities

During the experiment, we collected thermal videos and electrodermal activity (EDA) physiological parameters. The following devices were utilized for the data collection:

Thermal Imaging: The Seek Thermal CompactPRO thermal camera[1] [57] was used to capture thermal video data (i.e., sequence of thermal frames). Thermal frames were captured with a 240×320 pixel resolution, a 32-degree field of view, and at 5 frames per second (fps). We use libseek_thermal [110] API for data collection. Though the thermal camera can capture 10 fps, our observation showed that at 5 fps, frame rates are the most stable.

The Empatica E4 Wristband: The Empatica E4 Wristband[2] is a wearable device designed to monitor physiological signals and gather data about an individual's physical and emotional well-being. E4 wristband encompasses an EDA sensor and a PPG (Photoplethysmography) sensor, where EDA data is collected at a sampling rate of 4 fps, while the PPG data is captured at 64 fps.

4.2 Data-Collection Experimental Procedure

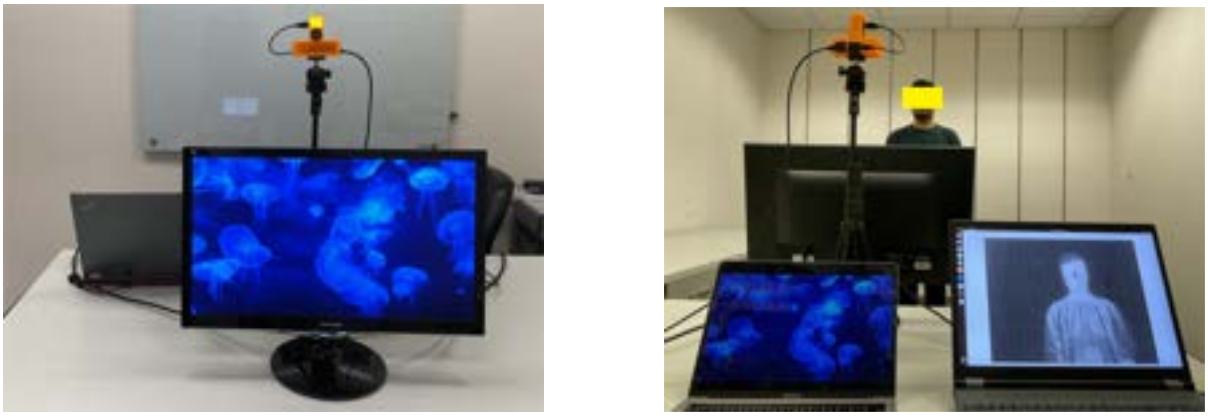
This section discusses the experimental procedure we followed during the data collection session. Upon arrival, the participants were given time to read and sign the consent form. Next, participants were asked to stand in front of a computer screen and a Seek Thermal CompactPRO camera. Additionally, they wore an Empatica E4 wristband on their left hand, which recorded their EDA responses during the data collection. Figure 2 illustrates the data collection setup in the laboratory.

Distances: Three distance lines were marked from the thermal camera at 5, 7, and 9 feet. Each participant was randomly asked to stand and encouraged to limit their movement within $1 - 2$ feet behind (i.e., farther from the sensor) one of these lines. In total, 10 participants stood at the 5 feet line, 12 participants at the 7 feet line, and 10 participants at the 9 feet line.

Ambient Temperature: Data collection was performed in indoor rooms of the X University over two years, across all seasons. However, indoor temperatures were regulated. During the heating season (September 15 - May 15), the AC was set to 68 degrees Fahrenheit; during the cooling season (May 16 - September 14), it was set to 76 degrees Fahrenheit.

Study Protocol: Thermal and EDA physiological response data were collected for non-stress and stressful conditions. Notably, this study's data collection protocol did not adhere to a 1-to-1 non-stress vs. stress challenge

[1] https://www.thermal.com/compact-series.html
[2] https://www.empatica.com/research/e4/



(a) Frontend (b) Backend

Fig. 2. Data collection setup

design. Instead, it followed a protocol similar to the Trier Social Stress Test (TSST) [34, 46], where one or more non-stress-inducing baselines are used for comparison with stress challenges. For instance, Iqbal et al. 2022's [46] protocol had a single non-stress-inducing baseline task followed by three stress-inducing tasks in a fixed sequence. Similarly, in this paper's protocol, participants performed two non-stress tasks first, followed by four stress-inducing tasks known to elicit physiological responses [11, 34, 46]. While two non-stress tasks enable establishing baselines from various conditions and participants' activities, each stress-inducing task presented unique stimuli to the participants, eliciting stress responses across various conditions and activities, as discussed below. The collected data from this protocol enables the development of a non-stress vs. stress detection approach applicable across various conditions.

Moreover, human physiological responses, including skin temperature changes, are not momentary concerning the onset of a stressor; rather, they may persist for several minutes [42]. Following the literature [11, 34, 46], to avoid any bias from residual stress effects, non-stress-inducing tasks are conducted initially, and later, a fixed sequence of stress-inducing tasks are performed sequentially. Additionally, as identified in [34], no interfering activities, such as questionnaires, occurred at least 15 mins before introducing the four stress-inducing tasks.

The study protocol is presented in Figure 3. On average, the data collection session was 15 minutes incorporating the gaps between the tasks. Participants self-reported their subjective stress levels on a scale from 0 (no stress) to 5 (extreme stress). Participants self-reported their subjective stress levels every 30 seconds during the watching calm video and stress-inducing video tasks. However, for tasks involving counting task, preparing a song, playing a number game, and recalling a negative memory, self-reporting was only performed at the end of each task. This approach was chosen to prevent potential disruption during task execution, which could affect the stressor's effectiveness. The details of the (1) Non-Stress Inducing and (2) Stress-inducing tasks are below.

Fig. 3. Study Protocol

Non-Stress Inducing Tasks: The participants performed two tasks in the non-stressful condition data collection phase.

(1) **Watching a Calm video:** Participants watched a 3 min video with jellyfish floating in the ocean. Recent stress and behavioral research studies [104, 108] have used similar videos to establish a non-stress-inducing measurement baseline. The self-reporting mean score was 0.73 during this task.
(2) **Counting task:** We asked the participants to slowly count from 0 to 59. This task was designed to reflect a non-stressful speaking condition. This task took, on average, 1 min for each participant. At the end of this task, the self-reporting mean score was 0.83.

Stress Inducing Tasks: The participants performed four stress-inducing tasks in this stressful condition data collection phase. The tasks were designed based on prior studies [11, 34, 100] in psychology and behavioral science.

(1) **Passive stress induce video**: Participants watched four stress-inducing video clips (for a total 3min) from the emotional stimuli database [100], which contains movie clips with labels: stressful, scary, fearful, and disgust, and are annotated and ranked by 50 film experts and 364 volunteers. During this task, the self-reporting mean score was 2.99.
(2) **Sing-a-Song Stress Test (SSST)**: During the SSST [11] task, participants were asked to prepare a song in 30 seconds without any prior notification in the presence of the task by the experiment coordinators. Subsequently, they sing a song for up to 30 seconds. It's important to note that this study used only the 30-second preparation phase data, excluding any data during the actual singing. This is due to research indicating that the SSST task, like song preparation, induces stress through social evaluation and uncertainty in the confederate's reaction to the participant's performance [24]. At the end of this task, the self-reporting mean score was 1.42.
(3) **Trier Stress Task (TST)**. This task follows the TST task [58], where the participants were given a surprise arithmetic task to count backward from a large number by '17'. For example, if the starting number is 1000, the participant should say: 983, 966, 949,..., etc. Every time the participants made a mistake or took longer, they were asked to start from the beginning. Studies [34, 40] have reported that TST is the gold standard protocol that leads to a reliable high-stress response. The task took, on average, about 3 mins for each participant. At the end of this task, the self-reporting mean score was 3.33.
(4) **Recalling a Bad memory:** According to literature [22], when we reminisce about negative events, our bodies respond as if we are experiencing those events again, activating the fight or flight response and releasing stress hormones such as cortisol and adrenaline. This can lead to high-stress response [60]. The task took, on average, about 2 mins for each participant. At the end of this task, the self-reporting mean score was 1.8.

4.3 Validating Stress-Response due to the Stress-Inducing Tasks

To verify induced stress through the stress-inducing tasks, we assessed Heart Rate Variability (HRV) using Empatica E4 wristband data. Literature [13, 54, 99, 111] link HRV changes to stress, often tied to reduced parasympathetic activity, seen as decreased High Frequency (HF) and increased Low Frequency (LF). We compute LF/HF ratio by dividing LF power by HF power, which rises under stress [54]. Following [13]; we extract photoplethysmogram (PPG) features from Empatica E4, with 3-minute windows and 1-second steps, aligning with recommended HRV analysis window sizes [73].

First, we separate the non-stress-inducing (i.e., 3-min windows belonging to the first two tasks in Figure 3) and stress-inducing task windows (i.e., 3-min windows belonging to the last four tasks in Figure 3), then filter all using a Chebyshev II order-4 filter (20 dB stopband attenuation, $0.5 - 5$ Hz passband). PPG signals become heartbeat intervals, removing outliers beyond $500 - 1200$ ms (heart rates $50 - 120$ bpm) and filling gaps with linear interpolation. Finally, we extract LF/HF HRV values from processed features.

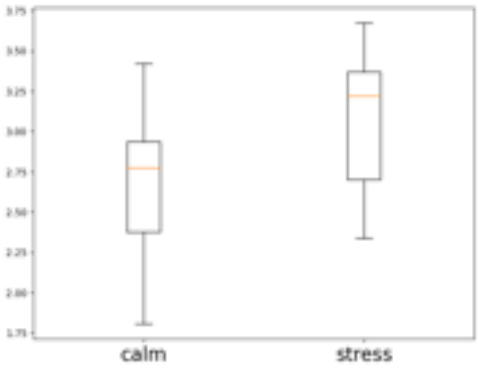


Fig. 4. LF/HF ratio during non-stress vs. stress-inducing tasks

Figure 4 shows the LF/HF ratio during the non-stress and stress-inducing tasks. We applied a one-way ANOVA, which yielded a statistically significant effect of stress on heart rate variability, i.e., LF/HF ratio (p-value=0.0077). Consistent with findings by [54], an elevated LF/HF ratio is noticeable during stress-inducing tasks, indicating heightened participant stress levels resulting from the introduced challenges through these four tasks.

4.4 Data Preprocessing

Notably, throughout our assessment of the *ThermaStrain*, we refrained from excluding any data segments based on self-reported information (as described in Section 4.2) or stress validation through HRV (discussed in Section 4.3). This decision was made because the physiological stress response may persist for several minutes concerning the onset of a stressor [42], and in each of the stress-inducing tasks, both self-reports and HRV analysis indicated heightened stress levels, confirming the efficacy of the stressors. Given that our experimental protocol entails stress-inducing tasks lasting between 30 seconds and 3 minutes each, assuming a participant is stressed, only a portion of that duration may introduce bias.

In this section, we discuss our data preprocessing steps for the raw thermal and EDA data.

Normalizing the EDA data: Research shows that individuals from different populations may exhibit different levels of skin conductance due to various factors such as genetics, skin thickness, and environmental factors. Therefore, following previous research [10], z-score normalization was applied to each participant to control for individual differences in EDA level.

Extraction of stress event detection windows: A window size of 5 seconds with 2 seconds of overlap was used for real-time stress detection. We determined this window size empirically (through grid search), aiming to balance high-stress assessment efficacy with the real-time usability of the sensing system. Larger window sizes yielded similar effectiveness, while smaller sizes compromised stress-assessment performance. Notably, this aligns with the discussion in Section 2. Given that thermal response latency is under 5 seconds, and the latency gap between EDA and thermal is approximately 2 seconds, a 5-second detection window can capture the physiological stress response at the onset of the stressor and facilitate knowledge transfer from EDA to Thermal during training.

During data collection, each thermal and EDA data point was marked with an absolute global time to facilitate synchronization between modalities. We synchronized the thermal and EDA data by selecting 5 seconds of thermal data and retrieving simultaneous EDA data according to the absolute global timestamp. After pre-processing, 5 seconds of thermal data had a dimension of [25 × 1 × 240 × 320] (an individual thermal frame was of the shape [1×240×320]), and the EDA data had a dimension of [5 × 4].

Human body detection: The human body constitutes a fraction of the thermal frames. Since only the human body thermal information is pertinent to the stress, we applied a human body segmentation (i.e., body-region identification) algorithm named DetectorRS [91] on thermal frames. We pre-trained the DetectorRS model on Microsoft COCO dataset [70] and evaluated its performance on our manually labeled (with pixel-wise body area and bounding box labels) thermal dataset. We used IOU as the evaluation metrics [95] that represent the ratio of overlap vs. union of the predicted and ground truth image segmentation regions. The DetectorRS body segmentation model achieves an average of 85.02% IOU score, which is reasonably high. After identifying the body region, the thermal frame pixel values outside the human body segmentation mask are zeroed out. Finally, a window-wise z-score normalization is applied to the thermal images. In all of the evaluations, background masked-out thermal frames are utilized during training and testing.

Such background masking enables *ThermaStrain* to be generalizable and readily deployable in unknown scenarios. E.g., as discussed in Section 8.2, such masking allows for identifying high stress in scenarios when multiple individuals are present in front of the thermal camera.

5 PROPOSED CO-TEACHING APPROACH

5.1 Problem Statement

Given a sequence of thermal frames, i.e., thermal video $t \in T$, where $t = (t_1, ..., t_k)$ and synchronized EDA values, $e \in E$, where $e = (e_1, ..., e_l)$, our goal is to train a model that can predict $y \in Y$, where y =*('stress' or 'non-stress')*, from only thermal video t without requiring the EDA values at inference time. Since thermal video and EDA sampling rates are not necessarily the same, for a fixed stress detection window, $k \neq l$.

5.2 Approach Overview

This section presents a novel co-teaching approach named *ThermaStrain*. Since EDA is a strong indicator of stress [102], the *ThermaStrain* approach simultaneously learns separate stress-indicative embeddings from EDA and thermal video; however, enforcing them to be similar for the same objective, inferring stress vs. non-stress class. *Such enforcement enables the extraction of knowledge from thermal video similar to stress-indicative EDA physiological parameters alongside other thermal attributes indicative of stress, resulting in a higher stress inference performance.*

The *ThermaStrain* model, shown in Figure 5, comprises three neural network modules. A thermal encoder F_T is used to map an input thermal video t to thermal embedding z_t; an EDA encoder F_E is utilized to map the synchronized EDA sequence e to EDA embedding z_e; And a classifier module F_C that predicts the corresponding inference y taking z_t and z_e separately.

5.2.1 During training. Both the thermal videos and corresponding synchronized EDA values are available during training. Meaning, the training dataset $D_{train} = (t, e, y)$, where $t \in T$, $e \in E$, and $y \in Y$. The stress vs. non-stress output is inferred through two streams. The thermal embedding z_t and EDA embedding z_e are generated separately. Notably, the generated z_t and z_e embeddings have the same dimension d. The classifier module F_C infers y^t and y^e by taking the z_t and z_e separately as follows:

$$\begin{aligned} y^t &= p(y|t, F_T, F_C) = p(y|z_t, F_C)p(z_t|t, F_T) \\ y^e &= p(y|e, F_E, F_C) = p(y|z_e, F_C)p(z_e|e, F_E) \end{aligned} \tag{2}$$

Task losses: One of the training objectives is to teach the model to infer y^t and y^e as close as the target output y. To attain the objective, the proposed approach introduces two task losses l_t and l_e, which are cross-entropy losses between the target output y and the generated inferences y^t and y^e through the two streams, thermal, and EDA. *Minimizing l_t and l_e ensures better embeddings z_t and z_e extraction encompassing the respective mortality's stress-indicative markers.*

$$l_e(y^e) = -logp(y|e, F_E)$$
$$l_t(y^t) = -logp(y|t, F_T) \tag{3}$$

Similarity loss l_s : Since the *ThermaStrain* approach also aims to *learn the extraction of knowledge from thermal video similar to the knowledge from EDA information that is highly predictive of stress, it is important to enforce the embeddings* z_t *and* z_e *to be similar.* Hence, *ThermaStrain* introduces a similarity loss l_s to maximize the joint likelihood of z_t and z_e during training. Here we use the mean square error to measure the similarity of embeddings.

$$l_s(z_t, z_e) = \sum(z_t^2 - z_e^2) \tag{4}$$

Consistency loss l_c : Another training objective is to encourage consistency between inferences y^t and y^e. Considering that with the utilization of similarity loss l_s during training, the thermal and EDA embeddings z_t and z_e would be similar; however, not the same, classifier module F_C may generate mismatched y^t and y^e. This may result in inferior performance during testing when the EDA sequence e would not be available. Hence, to enforce consistency between y^t and y^e, We define a consistency loss l_c.

$$l_c(y^t, y^e) = y^t log\frac{y^t}{y^e} = p(y|t, F_T, F_C)log\frac{p(y|t, F_T, F_C)}{p(y|t, F_E, F_C)} \tag{5}$$

Here, Equation 5 is the Kullback-Leibler divergence between the two conditional distributions y^t and y^e.

Overall Training Loss L: The overall optimization objective, i.e., overall training loss L of the *ThermaStrain* approach, is finally defined as a weighted sum of the two task losses, similarity loss, and consistency loss:

$$L = l_t + l_e + \alpha l_s + \beta l_c \tag{6}$$

Where α and β are hyperparameters that control the weight of co-teaching and consistency objectives during training.

5.2.2 During Test/Evaluation. Only the thermal videos are available during testing or evaluation. Meaning, the test dataset $D_{test} = (t, y)$, where $t \in T$, and $y \in Y$. As shown in Figure 5, during testing, the stress vs. non-stress output is inferred through the thermal stream, generating inference $\hat{y}$ taking t as input using the equation below:

$$\hat{y} = p(y|t, F_T, F_C) = p(y|z_t, F_C)p(z_t|t, F_T) \tag{7}$$

5.3 Discussion of the Modules

The modules of the *ThermaStrain* approach are discussed below.

5.3.1 Thermal Encoder F_T. It takes a thermal video in the form of a sequence of thermal frames, $t = (t_1, ..., t_k)$ as input, and generates an aggregated embedding z_t comprising stress indicative thermal markers/information from each frame, and the temporal thermal attributes depicted through the frame sequence. The module comprises a ResNet followed by a Transformer network.

As shown in Figure 5, ResNet takes each frame t_i to generate a framewise embedding z_t^i, representing the stress indicative thermal information from the respective frame. Later, the Transformer takes all the framewise embeddings $(z_t^1,\ldots,z_t^T)$ to aggregate the temporal information and generate the thermal video embedding z_t.

5.3.2 EDA Encoder F_E. The input EDA values e within a detection window have a [5×4] dimension representation. First, we extract the 6 features from EDA values, including mean, min, max, median, variability, and standard deviation. Then these six features was fed into two linear layers followed by a ReLU activation function to generate the EDA embeddings z_e.

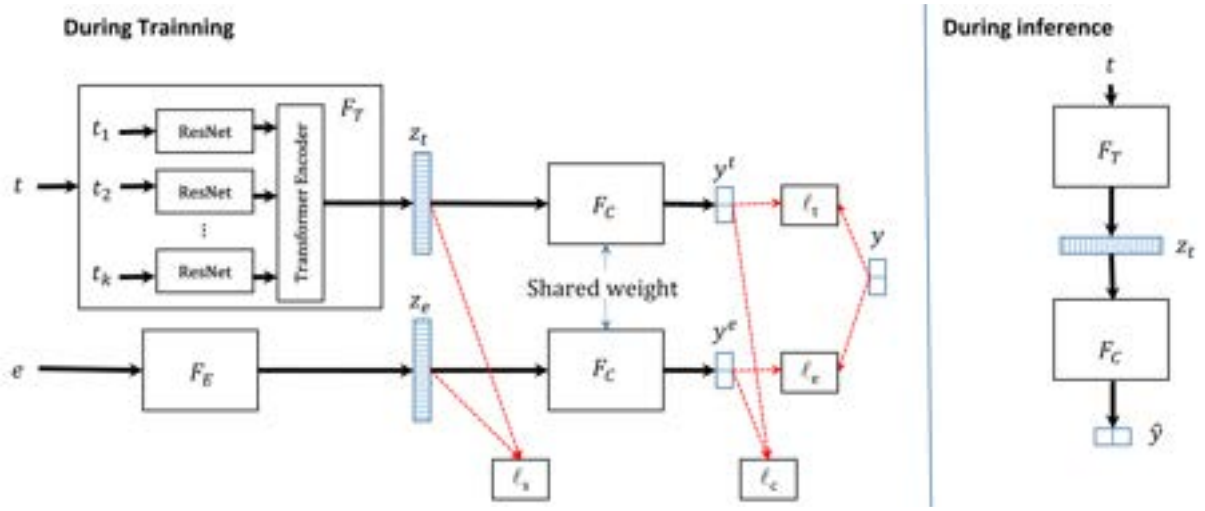


Fig. 5. The framework of *ThermaStrain* approach.

5.3.3 Classifier F_C. In the *ThermaStrain* model, the F_C module is represented by a simple network comprising linear layers with a ReLU activation function. Notably, the generated thermal and EDA embeddings z_t and z_e have the same dimension d. During training, the F_C takes both the embeddings separately to predict y^t and y^e, respectively. Finally during test/evaluation, F_C predicts the stress vs. non-stress inference $\hat{y}$ during the test or evaluation.

6 EVALUATIONS

This section discusses the efficiency and applicability of *ThermaStrain* by investigating some key questions. The presented network parameter configurations were optimized by performing a grid search of the possible parameter values. Presented evaluation results are end-to-end, incorporating the inaccuracy due to pre-processing errors. Finally, evaluations are presented with metrics: sensitivity, specificity, accuracy (%), and F1 scores.

Evaluation Dataset-split: For each of the evaluations, we followed ***the person-disjoint hold-out method*** [15]. Our collected data includes 32 sessions (each with a different participant). In this study, we performed a 5-fold evaluation where-in each fold, we split the dataset into validation (5 sessions) and training set (rest of the sessions). Training and validation sets were disjoint concerning participants and sessions. Our dataset is imbalanced, having more stress samples than non-stress. Therefore, in each fold, we performed under-sampling on the stress samples to make the dataset balanced.

6.1 Implementation of the Presented Approach

As discussed in Section 5.2, presented approach has three modules, discussed below:

- *The Thermal Encoder*: Comprises a ResNet and a Transformer Encoder. The ResNet starts with a 7×7 convolutional layer, followed by three residual blocks. Each residual block includes two convolutional layers, followed by batch normalization and ReLU activation. At the end of the ResNet, an adaptive pooling layer pools the feature map into a size of 2×2. Finally, we set the channel of the last convolutional layer as 64, resulting in an output embedding of shape $2 \times 2 \times 64$, which is 256 dimensions. The embedding of all frames is then fed into a transformer encoder to aggregate information over time. After the transformer,

we perform mean pooling on the sequential dimension and use the resulting embeddings for downstream tasks.

- *The EDA Encoder*: Considering the input EDA data has a simple feature dimension of 6. We design a simple EDA encoder that comprises two linear layers with ReLU activation and dropout.
- *The Classifier*: The classifier module comprises two linear layers with a ReLU activation function and dropout.

6.2 Optimization

The weights α and β in Equation 6, and the learning rate are hyper-parameters that were identified through the python toolkit Optuna[2]. It uses a Bayesian Optimization algorithm called Tree-Structured Parzen estimator to identify the optimum set of values.

6.3 Comparison of Co-Teaching with Uni- and Multi-Modal Approaches

This section compares the *ThermaStrain* approach with uni-modal and multi-modal approaches leveraging the modalities in hand, thermal video, and EDA. To ensure a fair comparison among the approaches being compared, we set the complexity of each component to be the same, including the number of neurons in linear layers, the number of layers in the classifier, and the number of kernels in the CNN layers. We then use Optuna, a hyperparameter optimization framework, to comprehensively evaluate hyperparameters such as the learning rate to determine the optimal settings. Finally, we present the best results obtained for each model.

Implementations: The thermal baseline takes only a 5-second thermal video as input and predicts stress. It uses ResNet as a feature extractor, a transformer to aggregate information over time, and a multi-layer Perceptron classifier to make the classification.

The EDA baseline takes 5 seconds of EDA data and predicts stress. It uses a multi-layer perceptron followed by a transformer to extract features and a multi-layer perceptron classifier to make the classification.

The multimodal baseline shares a similar structure to our *ThermaStrain* implementation, with a thermal feature extractor, EDA encoder, and classifier. The only difference is that, instead of enforcing the z_T and z_E embeddings to be similar, classifier taking each of them separately, the classifier takes the concatenated embedding of z_T and z_E to make inferences.

Evaluation Result Discussion: Table 1 presents the results of our stress vs. non-stress binary classification evaluation. The *ThermaStrain* approach achieved an accuracy of 83.17% and an F1-score of 0.8293. In comparison, the uni-modal thermal baseline model only achieved 76.2% accuracy and 0.7592 F1 scores. The *ThermaStrain* model outperforms the thermal baseline model by over 9%.

The EDA baseline and multi-modality model also achieved 0.8568 and 0.8897 F1 scores, respectively. These scores are higher than our thermal baseline and *ThermaStrain* model.

The evaluation coherent with EDA is a strong indicator of stress. The co-teaching approach successfully extracts EDA-relevant embedding, i.e., skin conductance-relevant information from thermal video, resulting in a significant performance improvement over the uni-modal thermal baseline. However, such extraction is lossy; hence co-teaching still cannot outperform multi-modality or EDA-based stress assessment approaches. Notably, compared to these approaches, EDA is not required by the *ThermaStrain* approach at the inference time, enabling contactless deployment, i.e., less obtrusive sensing.

6.4 Generalizability Evaluation

A generalizable stress-sensing solution needs to be robust and perform similarly in previously unseen stress conditions, meaning in the stressful situations that were not present in the training dataset.

Model	Sensitivity	Specificity	Accuracy	F1 score
Thermal baseline	0.8397	0.6663	76.2%	0.7592
EDA baseline	0.8559	0.8575	85.71%	0.8568
ThermaStrain	0.8911	0.7598	83.17%	0.8293
Multi-modality baseline	0.9202	0.8541	**89.13%**	**0.8897**

Table 1. Evaluation of *Co-teaching* compared with Uni- and Multi-Modal Approaches

To evaluate the generalizability of the *ThermaStrain* approach in unseen stress conditions, we performed a stress-task disjoint evaluation over the 5-folds (discussed in Section 6). As mentioned in Section 4.2, each participant performed four distinct stress-inducing tasks in each data collection session, simulating four stress conditions. In this evaluation, during training, only the 'Passive stress induce video' and 'TST' task data were used as stress samples, while during evaluation, only the 'SSST' and 'recalling bad memories' task data were used.

Like the previous section, *ThermaStrain*'s performance is compared with EDA and thermal video-based uni-modal and multi-modal approaches; the results are shown in table 2.

The uni-modal thermal baseline's performance decreased by 3% in accuracy and F1 scores compared to the baseline evaluation in table 1. In contrast, *ThermaStrain* model's performance decreased by 2%, but it is still significantly better than the thermal baseline.

The EDA and multi-modality baselines achieved even higher accuracy. These evaluations indicate that during the 'SSST' and 'recalling bad memories' tasks, EDA is an even stronger indicator of stress than thermal modality.

It is important to note that, according to our study protocol discussed in Section 4.2, the stress responses for each task are not completely disjoint. This is due to the potential partial influence of residual physiological stress responses from stress-inducing tasks on one another. Nonetheless, the evaluation in this section demonstrates the relatively greater generalizability of the *ThermaStrain* approach compared to the unimodal thermal baseline, thereby showcasing its improved utility.

Model	Sensitivity	Specificity	Accuracy	F1 score
Thermal baseline	0.8373	0.6666	73.24%	0.7289
EDA baseline	0.9107	0.8815	**89.34%**	**0.8939**
ThermaStrain	0.9028	0.7595	81.36%	0.8107
Multi-modality baseline	0.9586	0.8517	**89.34%**	0.8919

Table 2. Evaluate on unknown (during training) stress tasks

6.5 Comparsion with Other Co-teaching Baselines

This section compares *ThermaStrain* approach with the existing co-teaching baselines. Following the state-of-the-art literature [29, 62, 68], we implemented three co-teaching approaches as baselines. Since none of them have leveraged thermal and EDA modalities, we followed their model structure and design but made necessary changes to fit our dataset.

The results are shown in table 3. As discussed in Section 3, the multi-task learning approach [29] has multiple classifiers that fit different missing modality scenarios. In StressNet [62], thermal data was used to predict the EDA modality and then used the predicted EDA to predict stress. In the vision hallucination model [68], there is a hallucination network that mimics the EDA embedding. During inference, pseudo-embedding replaces the EDA embedding and concatenates with the thermal-independent embedding.

As shown in table 3, all models achieved lower performance than *ThermaStrain*. The reason is that the multi-task learning approach doesn't have similarity loss and consistent loss that force each modality to learn joint patterns. The StressNet only takes the reconstructed EDA modality to predict stress, which loses some independent information about the thermal modality. *The inferior performance of StressNet further emphasizes the impact of*

presented co-teaching approach rather than just simulating physiological parameters from thermal sensing and using the simulated physiological features to assess stress. Finally, the vision hallucination model is too complex, leading to overfitting in our limited dataset.

Model	Sensitivity	Specificity	Accuracy	F1 score
ThermaStrain	0.8911	0.7598	**83.17%**	**0.8293**
multi-task learning	0.8145	0.7595	79.05%	0.7891
StressNet	0.8443	0.6930	77.48%	0.7739
Visual hallucination	0.8268	0.7317	78.5%	0.7826

Table 3. Comparsion with other co-teaching approaches

6.6 Ablation Study

Table 4 presents the results of the ablation study, where various components of *ThermaStrain* are modified while keeping the rest of the network constant. We discuss the evaluation results and observations below:

Using Central Moment Discrepancy (CMD) as l_s: Both Mean Square Error (MSE) and CMD loss are popular distance metrics that measure the discrepancy between the distribution of two representations. We evaluate them thoroughly. As shown in table 4, the MSE achieves better performance. Therefore, we choose MSE as our l_s.

Not using l_c: As there is a similarity loss l_s that forces the two embeddings z_T and z_E to be similar to each other, it may be questioned whether we need the consistency loss l_c. Hence, we evaluated not using the l_c in the *ThermaStrain* implementation. The results show a drastic performance drop when we remove the consistency loss, demonstrating the importance of l_c in encouraging consistency between inferences y_t and y_e, which leads to better performance while EDA modality, consequently, y_e is unavailable during evaluation.

Use Vision Transformer to replace ResNet: The Vision Transformer [25] is a transformer-based image feature extractor and outperforms ResNet-based models in RGB image-based literature. We attempt to use the Vision Transformer as our feature extractor for each frame. However, accuracy and F1 scores decrease by approximately 5%. Considering the limited size of our dataset, the complex transformer-based feature extractor may lead to overfitting.

Model	Sensitivity	Specificity	Accuracy	F1 score
ThermaStrain	0.8911	0.7598	**83.17%**	**0.8293**
l_s *use Central Moment Discrepancy*	0.8208	0.7555	79.17%	0.7909
Not use l_c	0.7871	0.7608	77.64%	0.7753
Use Vision Transformer to replace ResNet	0.7646	0.7581	76.13%	0.7603

Table 4. Ablation study of our proposed model

6.7 Distance Evaluation

This section breaks down the performance of the *ThermaStrain* based on the distance between the participant and the thermal camera. It is observed that the model performs better for closer distances. Specifically, the model achieves the highest performance in the 5-7 ft range with an accuracy of 91.36% and an F1 score of 0.9126. However, the performance drops for larger distances, especially in the 9-11 ft range, where the accuracy is 72.58%, and the F1 score is 0.6902. The performance deterioration for larger distances can be attributed to each pixel's reduced quality of thermal energy perception and the reduction in the quantity of thermal pixels covering the important body regions at the longer ranges.

Model	Number of sessions	Sensitivity	Specificity	Accuracy	F1 score
5-7 ft	7	0.9304	0.8904	**91.36%**	**0.9126**
7-9 ft	8	0.8241	0.9016	87.81%	0.8634
9-11 ft	7	0.8826	0.5187	72.58%	0.6902

Table 5. Distance evaluation

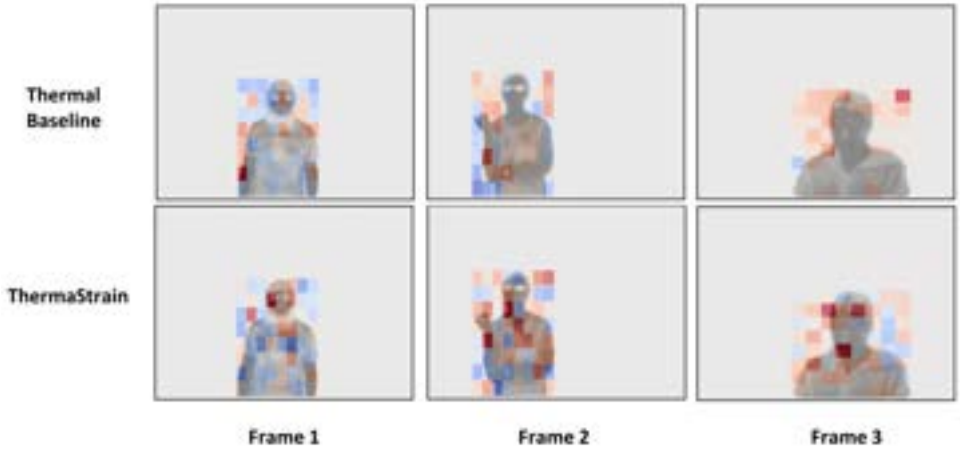


Fig. 6. The SHAP interpretation to Visualize Models' Information Extraction Efficacy.

6.8 Benchmarking for Real Time execution

To evaluate our approach's real-time executability, we performed a run-time evaluation on an Nvidia Jetson Nano. The program reads 5-second data at a time, analyzes and infers the class results, and waits for the next 5-second data. The binary classifiers take 0.324s to process one 5-second video window on Jetson Nano. The average CPU usage is 19.38%, and the average GPU usage is 9.64%. The average RAM usage is 1.85 GB. According to this evaluation, our presented approach is capable of real-time execution on a Jetson Nano module. Note that the times reported are when only the stress assessment program is running. Running additional programs will affect/change these times.

7 DISCUSSION ON *THERMASTRAIN*'S EFFICACY

This section further investigates the *ThermaStrain*'s capability in identifying effective thermal information extraction (Section 7.1) and how co-teaching enables better stress sensing solution development through loss landscape analysis (Section 7.2).

7.1 Inference Interpretation Discussion: Visualizing *ThermaStrain*'s Effective Information Extraction

Many studies have highlighted that stress-induced changes in temperature are primarily concentrated in the forehead, eyehole, and cheekbone regions [82, 88]. Therefore, these regions are more critical for improving the accuracy of stress detection models. To evaluate the developed models' capability in capturing information from the critical body regions, we utilized the KernalSHAP model-agnostic interpretation framework [72]. The SHAP values [103] indicate the contribution of each input attribute in driving the model inference closer or farther away from the true/correct inference. We divided the human body region into an 8×8 grid to compute the Shapley value for each grid.

Figure 6 shows generated explanations of the baseline and *ThermaStrain* classifiers' inference for three thermal frames belonging to three different individuals. The baseline model's Shapley value appears more normally distributed, indicating that it had to select information from the entire frame. In contrast, *ThermaStrain* effectively learns the critical regions to focus on, illustrated by having darker colors in the face, neck, and hand regions that are established as crucial stress-indicative body areas according to literature [82, 88].

This visualization demonstrates that *the EDA modality effectively guides the* ThermaStrain *model to extract better thermal embeddings.* This results in perceiving high-stress-indicative and physiologically relevant information by focusing on crucial visible body regions.

7.2 Loss Landscape Visualization: Understanding How Co-teaching Facilitates Effective Model

Our evaluation in Section 6.3 shows that the presented Co-teaching approach outperforms the uni-modal thermal sensing baseline approach. This section investigates how the co-teaching approach enables better performance.

Li et al. [67] showed that visualizing the loss landscape for neural network models provides a richer understanding of how the different approaches' design choices influence the optimization of the loss function. We used the *loss-landscapes* library [75] to generate the 3D loss landscape plots of *ThermaStrain* and the uni-modal thermal baseline model, as shown in Figure 7. A detailed discussion on the plot generation is in Appendix A.1.

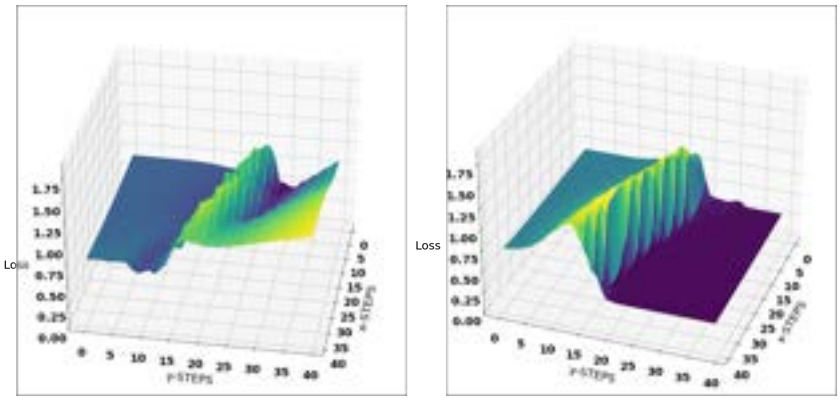


Fig. 7. 3D Loss Landscape Plot Comparison (Baseline is in the left vs the *ThermaStrain* is in the right)

Several prior works [17, 43, 52, 67] investigating the loss landscapes to understand the ability of neural networks to optimize better (i.e., obtaining better performance) emphasized that the 'flatness' of the loss landscape is a property of interest. Hochreiter et al. [43] define the 'flatness' of the loss landscape as the region around the minima where the loss remains low. Literature [43, 50, 67] suggests that the model with flatter loss surface optimizes better, i.e., effectively identifies the minima in the loss space, hence achieves better performance.

As shown in Figure 7, the loss landscape is significantly 'flatter' in the blue regions (near minima) for the *ThermaStrain* than the baseline approach. Meaning *co-teaching enables easier propagation through the loss landscape and identification of minima, resulting in a more effective stress sensing performance* by the *ThermaStrain* model.

8 DISCUSSION ON APPLICATIONS AND DEPLOYMENT OF THE *THERMASTRAIN*

This section discusses the scenarios where *ThermaStrain* will be more effective than wearable-based stress sensing solutions (Discussed in Section 8.1). Additionally, it expounds upon the deployment challenges associated with deploying *ThermaStrain* in real-world settings (Discussed in Section 8.2). This discussion encompasses scenarios involving multiple individuals, constrained visualization, and deployment conditions, camera angles, distances,

backgrounds, participant's postures, etc., that were not encountered during the training phase. Finally, the Section 8.3 discusses the ethical and practical considerations for real-world deployment of *ThermaStrain*.

8.1 Application Scenarios of *ThermaStrain*

As shown in Section 6.3, while *ThermaStrain* outperforms the thermal-video-based state-of-the-art solutions, its efficacy is relatively lower than the EDA-based stress sensing solutions. Nevertheless, contactless thermal video-based stress detection presents distinct applications and deployment possibilities that are challenging to achieve through EDA or other wearable-based solutions. Two such example scenarios are discussed below.

8.1.1 Application in Smart Health. While wearable sensors find extensive application in healthcare monitoring scenarios [8], contactless sensors present distinct advantages over wearables in specific situations.

For instance, the non-intrusive characteristics of contactless sensors render them especially suitable for assessing stress in vulnerable populations, such as elderly individuals who may have impaired memory function, as observed in cases of Dementia [28, 59, 98]. Wearable sensors necessitate patients to wear or carry battery-powered devices, which can lead to discomfort and inconvenience due to frequent recharging of batteries [115]. In the case of Dementia, patients might forget to wear or charge these devices. Additionally, wearable sensors can pose risks to the safety of elderly individuals, e.g., a recent incident involved an elderly woman strangled by her fall detection pendant [3].

Consequently, contactless sensing solutions present an effective continuous stress assessment alternative. While RGB video-based solutions are already making their way into commercial use for monitoring elderly health [97], privacy concerns limit their widespread adaptation [124]. Thermal imaging offers relatively higher privacy protection compared to RGB-based alternatives. For instance, unlike RGB imaging, *ThermaStrain* ensures contextual or environmental information protection by not capturing non-human body content [18, 35]. Hence, *ThermaStrain* holds the potential as a highly suitable option for such vulnerable populations. This approach requires no active participation from patients, such as device recharging or wearing, and enhances safety due to its passive contactless operation.

8.1.2 Application in Smart Work-Place. Stress monitoring in smart workplaces, whether in manufacturing contexts [66] or smart offices [5], is crucial for safeguarding employee well-being, optimizing productivity, enhancing workplace safety, reducing staff turnover, and fostering a positive work environment.

Presently, prevalent techniques often involve the measurement of EDA through disc electrodes [9] or through the utilization of wearable devices like the Empatica E4 or the Apple Watch [7]. Positioning disc electrodes at the most sensitive bodily sites, such as the feet or fingers [109], can impose significant inconvenience on users or may even be impractical, such as in office settings where hands are engaged in typing or other activities. Additionally, since workplaces involve multiple individuals, using wearables can incur substantial costs. Furthermore, some individuals find it uncomfortable to wear such wearable devices for extended durations consistently [48].

Hence, contactless but relatively privacy preserving *ThermaStrain* can be a suitable alternative capable of simultaneously assessing stress in multiple individuals at a relatively affordable cost. Notably, stress assessments aimed at quantifying employee well-being within smart workplaces [64, 74] often occur in an aggregated manner rather than being conducted in real-time, such as on a per-minute basis. This characteristic ensures that the efficacy of the use case remains unaffected, even in scenarios where employees might be momentarily obscured due to occlusion. Significantly, thermal camera-based solutions are already being incorporated into workplaces for tasks like employee health screening [16] and security measures [101]. This trend paves the path for smoother integration of thermal-video-based stress assessment solutions like *ThermaStrain* into smart workplaces.

[3] https://www.huffpost.com/entry/medical-necklace-strangles-woman_n_56d75817e4b0871f60edbb47

8.2 Deployment Procedure of *ThermaStrain* in Real-world Scenarios

Three challenges require attention for the successful practical implementation of *ThermaStrain*.

(1) Achieving effective body segmentation and accurately identifying segments corresponding to the target users undergoing stress assessment in scenarios involving multiple individuals is crucial (Discusses in Section 8.2.1).
(2) Due to real-world occlusion scenarios in multi-person settings, only partial body segments might be accessible. In such cases, *ThermaStrain* must demonstrate superior performance compared to the thermal and co-teaching baselines outlined in Sections 6.3 and 6.5 when dealing with partial body segment information (Discussed in Section 8.2.2).
(3) *ThermaStrain* needs to maintain its stress-assessment performance while evaluating on distances, angles, indoor settings, and scenarios that are not present during its training (Discussed in Section 8.2.3).

Evaluation and discussion on these challenges are below:

8.2.1 Segmentation and Person-Identification in Multi-person Scenarios. We developed an integrated framework combining pre-trained *human body segmentation* and *re-identification* models to achieve simultaneous human body segmentation and identification. Initially, thermal frames are inputted into the *segmentation model* to generate pixel-wise human segmentation, producing disjoint object segments. Subsequently, these object segments are fed into the *human re-identification* model to assess if it belongs to one of the target individuals whose stress is being assessed.

As discussed in Section 4.4, for *human body segmentation*, we use the pre-trained DetectorRS [91] model on Microsoft COCO dataset [70], that identifies the human body regions in the thermal frame and masks all other parts of the background. We use pre-trained Omni-Scale Network (OSNet) [122] for *human re-identification*. The OSNet comprises a residual block composed of multiple convolutional feature streams, each detecting features at a certain scale. This enables OSNet to learn omni-scale feature learning. We take the pre-trained checkpoint provided by the author [121].

We conducted a *small evaluation* for a five-person indoor stress assessment scenario to assess the framework's effectiveness. Initially, we gathered a few seconds of data from each of the five participants separately for fine-tuning the human re-identification model. Subsequently, we conducted three sessions where varying subsets of the five individuals appeared simultaneously in front of the camera. Two sessions involved three different participants, while the remaining session had four participants. For the evaluation of body segmentation and person identification, the data were annotated by two graduate students, achieving an inter-rater reliability rate of over 94.91% (0.94+), specifically measured using Cohen's kappa statistic [78].

Frames featuring multiple individuals were processed by the DetectorRS human body segmentation model to extract object segments. We fine-tuned the Omni-Scale Network using the initial data collected individually, then evaluated the model's performance on the three multi-person sessions. The fine-tuned Omni-Scale Network achieved an impressive 97.29% accuracy in re-identifying participants during concurrent appearances.

Figure 8 illustrates instances of the integrated framework in action. When multiple individuals are present, the framework distinguishes and separates their respective body segments, detects corresponding identifications, and generates distinct frames for each identified body region along with the person's identification label. Each individual's specific frame, containing only their body region's thermal data, is inputted into the *ThermaStrain* model for stress assessment.

Given that in multi-person scenarios, only partial body segments might be accessible, the subsequent section delves into the discussion about the effectiveness of the *ThermaStrain* model, as well as different thermal and co-teaching baselines, when dealing with partially occluded body segments.

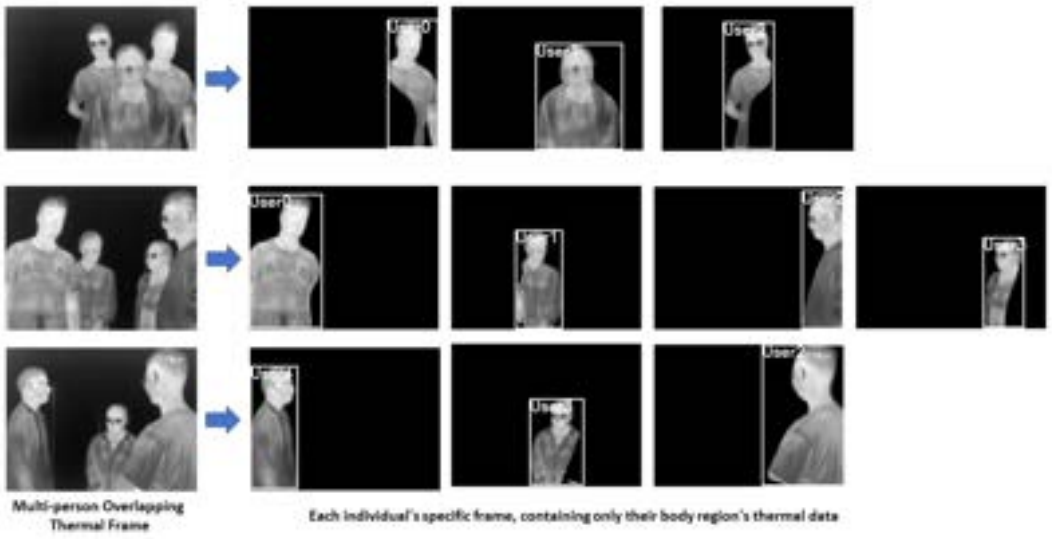


Fig. 8. Integrated multi-individual body segment and identification detection

8.2.2 Stress Assessment when Human Body Segment is Partially Masked. Occlusion presents a challenge in stress detection. This section discusses a pseudo-partially-masked-body-segment data augmentation to address this challenge. In this approach, during the training process, a portion of the participant's body is randomly masked with zeros. To facilitate this, a pre-trained multi-person human body part segmentation model named CDCL (Cross-Domain Complementary Learning) [69] is adapted, utilizing which we segment the human body into eight distinct sections. The CDCL recognizes pixel-wise human body part segmentations, such as the head, torso, upper arms, and forearms, from thermal frames. Subsequently, these segmentations were further refined into left and right divisions for each body part, resulting in eight subdivisions: the left face, the right face, the left torso, the right torso, the left upper arm, the right upper arm, the left forearm, and the right forearm.

We trained the *ThermaStrain* model alongside the baselines outlined in Sections 6.3 and 6.5, where there was a 23% chance of masking one out of the eight parts, a 23% chance of masking two parts, a 24% chance of masking three parts, and a 30% chance of keeping all parts unmasked. An example of pseudo one-body-part masked human-body segment frames, alongside the eight distinct body sections identified by CDCL from the original frame, is shown in Figure 9.

The evaluation outcomes for the pseudo-augmented trained *ThermaStrain* and the baselines are presented in Table 6. This evaluation encompasses scenarios where different body segments are masked/occluded. Comparing these results with those in Table 3, it becomes apparent that the training process involving significant partial noise causes a reduction of approximately 2% in *ThermaStrain*'s F1 score for 'without any masking body segments.' Nevertheless, the model sustains a satisfactory performance level across scenarios involving masking different body parts, consistently outperforming all the baselines. Notably, according to Table 6, masking the head and body exerts a more pronounced impact on performance than other parts. Additionally, masking any two body sections simultaneously reduces *ThermaStrain*'s accuracy and F1 score to on avg. 72% and 0.69 F1 scores, outperforming the thermal baseline (on avg. 68% accuracy and 0.65 F1 score) and best co-teaching baseline Hallucination (on avg. 69.7% accuracy and 0.66 F1 score).

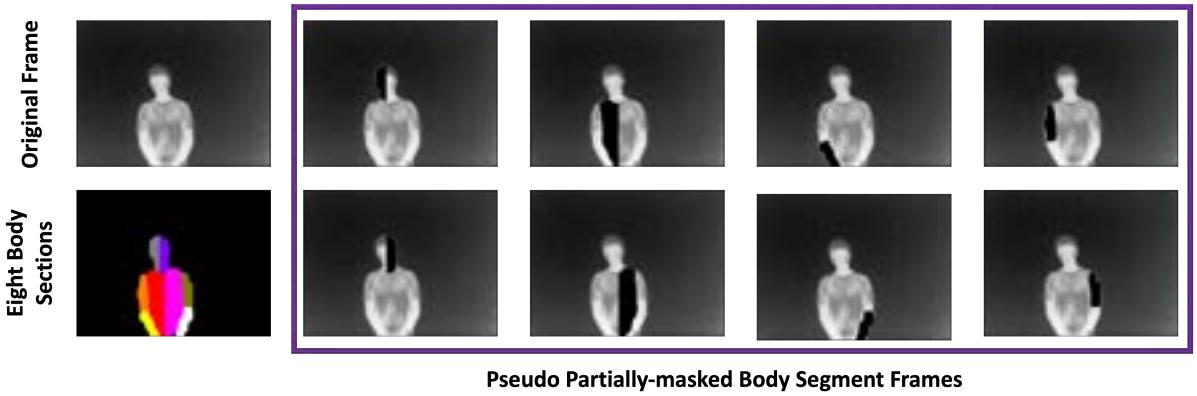


Fig. 9. Pseudo one-body-part masked human body segment frames.

More robust strategies are available to tackle the challenge of partial body masking due to occlusion. These methods include reconstructing the missing segments [114] and incorporating pyramid perception [**?**], which can potentially enhance the partial body stress sensing performance. However, considering this paper's primary focus is on co-teaching, the evaluation in this section was confined to pseudo-data augmentation. This evaluation demonstrates the viability of *ThermaStrain* in scenarios involving occlusion of partial body segments and its consistent outperformance compared to the baselines.

Finally, this section's evaluations and discussions demonstrate the *ThermaStrain*'s robustness against diverse occlusion scenarios compared to the baselines, thereby showcasing its improved utility in real-world settings.

	ThermaStrain		Thermal baseline		Multi-task		StressNet		Hallucination	
Metrics	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Without Masking	**0.81**	**0.80**	0.75	0.74	0.74	0.73	0.75	0.74	0.77	0.75
Masking Left Head	**0.79**	**0.78**	0.75	0.74	0.73	0.72	0.72	0.72	0.76	0.75
Masking Left Body	0.73	**0.73**	0.73	0.71	0.70	0.68	0.71	0.71	**0.74**	**0.73**
Masking Left Upper Arm	**0.79**	**0.78**	0.75	0.74	0.74	0.73	0.75	0.74	0.76	0.75
Masking Left Lower Arm	**0.80**	**0.79**	0.75	0.74	0.74	0.73	0.74	0.73	0.77	0.75
Masking Right Head	0.75	0.73	0.73	0.72	0.73	0.72	0.72	0.72	**0.77**	**0.76**
Masking Right Body	0.72	0.71	0.68	0.67	0.69	0.68	0.70	0.70	**0.73**	**0.72**
Masking Right Upper Arm	**0.78**	**0.77**	0.71	0.71	0.73	0.72	0.73	0.72	0.76	0.75
Masking Right Lower Arm	**0.79**	**0.79**	0.75	0.74	0.74	0.73	0.74	0.74	0.77	0.76

Table 6. Evaluation results of stress assessment on different masked sections of the Human Body.

8.2.3 Real Deployment Evaluation of ThermaStrain *on Unseen Distance, Camera-Angle, Indoor-setting and Scenario.* We deployed and evaluated the performance of the trained *ThermaStrain* model from Section 8.2.2 on a one-person workplace scenario in an indoor lab shown in Figure 10b. The participant in this study was a male who did not appear in the training set or validation set of the *ThermaStrain* model. In contrast to the data collection setup described in Section 4, the thermal camera was positioned at a forty-five-degree angle to the participant's left front and maintained a distance of about three feet from the participant. Also, the background environment setting (e.g., a whiteboard in the background) was different. The experiment spanned two consecutive days with a single individual: on the first day, the participant engaged in a LeetCode contest, simulating stress conditions, while on the following day, the participant relaxed by watching random YouTube videos of his choice. Each

(a) Thermal Frame Example

(b) Work-Place setup

Fig. 10. Work-Place Application Deployment Experiment

session lasted for approximately an hour. Data from the participant were captured using both the thermal camera and the Empatica E4 device.

Following the procedure outlined in Section 4.3, we calculated the LF/HF ratio of the participant's data collected with Empatica E4. The average LF/HF ratio for the first day was 1.71, which exceeded the ratio of 1.57 observed on the second day. This indicates higher stress was experienced by the participant on the first day. To statistically confirm this observed trend, we conducted a one-way ANOVA test, revealing a highly significant impact of stress (p-value=6.397e-07). No data segments were excluded from this section's analysis. We treated all data from the first day as stress instances and all data from the second day as non-stress instances.

The data collected from this experiment represents a real-deployment scenario for *ThermaStrain*, wherein the *thermal camera angle, distance, participant's posture (i.e., seated behind a desk instead of standing in Section 4), stress-inducing task conditions, indoor environment, and background* were unobserved during the model's training. Without any further adjustments (i.e., re-training), we evaluated the previously trained *ThermaStrain* model from Section 8.2.2 on this collected indoor one-person workplace scenario dataset. The evaluation yielded an accuracy of 84.59% and an F1 score of 0.8398, aligning with *ThermaStrain*'s performance in Table 1.

While a comprehensive study involving numerous participants, diverse scenarios, and varied indoor settings would be essential to establish *ThermaStrain*'s robustness to unseen data scenarios, this section's evaluation showcases its potential for real-world deployments across various applications.

8.3 Possible Ethical/Practical Considerations for Deployment

Ethical considerations for practical deployment scenarios are critical in human sensing applications. When deploying *ThermaStrain* in real-world scenarios, established strategies and protective measures from relevant literature [6, 37, 39, 77, 105] will be utilized to uphold the well-being, privacy, and rights of individuals.

Before gathering and processing thermal information, the application will secure informed consent from individuals whose data will be utilized. This ensures that participants understand data usage's purpose and potential consequences, allowing them to opt out. Clear communication regarding thermal information processing's aims, methods, and potential outcomes will foster trust and facilitate informed choices.

For those who opt out or do not pertain to the target group of individuals for stress assessment, their data will remain unused and unprocessed. As discussed in Section 8.2.1, it is possible to identify the body segments of the target users accurately. Our proposed deployment procedure (discussed in Sections 4.4 & 8.2) zeros out all other content of the thermal frames except the target user's body segment, hence other individuals' data will be

masked-out (i.e., zeroed out), and no processing will be performed on their information. Similarly, no information from the indoor environment will be processed, such as furniture, personal items, books, addresses, displayed documents, content within photo frames, and similar items, safeguarding against the leakage of environmental information and preserving privacy [18, 21].

Additionally, on-device computation offers superior security and privacy [116] compared to cloud-based alternatives. Our real-time evaluation in Section 6.8 verifies that the *ThermaStrain* approach can operate efficiently on resource-constrained edge platforms like Nvidia Jetson Nano, ensuring secure and privacy-preserving stress assessment in real-world settings.

The proliferation of thermal and image-based human-centric applications is driven by advancements in sensors and AI. Collaboration among experts spanning ethics, law, social sciences, and technology is imperative for a comprehensive ethical approach. While this paper adheres to prevailing ethical standards, future multidisciplinary collaboration endeavors will yield more well-rounded solutions for thermal human-centric sensing applications. Nevertheless, such endeavors remain beyond the scope of this paper.

9 DISCUSSION ON STUDY LIMITATIONS

This section discusses the study limitations and future research scopes of *ThermaStrain* approach.

Physiological Signals as Aiding Modality in Co-teaching: It is important to note that the presented solution showed that co-teaching enhances thermal stress sensing performance with the aid of EDA during training, a physiological sensing modality, during training. However, we also evaluated ECG and HR as aiding modalities. However, EDA outperformed others. Hence the presented paper includes only the EDA co-teaching solution and corresponding results. However, we cannot conclude inclusion of ECG or HR as a co-teaching aiding modality would not be beneficial. Exhaustive analysis with larger datasets and scenarios is needed to make such a conclusion, which was out of the scope of this paper.

In-the-wild Evaluation: A limitation of this study is that we specifically analyzed indoor data derived from laboratory environments. Given the primary focus on co-teaching, a comprehensive evaluation in real-world conditions was not within the study's scope. It is worth noting that Section 8.2 offers extensive analysis and discourse regarding the deployment of *ThermaStrain* in real-world multi-person scenarios, highlighting its superior viability in comparison to the baselines. However, numerous real-world factors remain unexamined. For instance, the impact of ambient temperature on thermal stress assessment was not assessed, as the data collection took place solely in temperature-controlled indoor settings with no recording of ambient temperatures for each session. Therefore, future work would benefit from sampling data from a broader range of in-the-wild situations to determine the boundaries of the *ThermaStrain* model's predictive validity.

Study Protocol: As detailed in Section 4.2, this study's data collection approach adhered to established literature in behavioral science and psychology [11, 34, 46]. For instance, to avoid any bias from residual stress effects, non-stress-inducing tasks were followed by stress-inducing tasks [11, 34, 46]. Additionally, as highlighted by [34], no interfering activities, like questionnaires, occurred at least 15 minutes before introducing the stress-inducing tasks. Our analysis with HRV on Section 4.3 confirms heightened participant stress levels during stress-inducing tasks, indicating the protocol's effectiveness. Nonetheless, a more comprehensive assessment, such as randomizing the order of the stress-inducing tasks, would unveil the most efficacious stress-inducing protocol. This falls within the domain of behavioral science/psychology research and is beyond the scope of this paper.

Age, Sex, and Demography: With respect to sex, our dataset was relatively balanced (12 male, 20 female). Our analysis showed that *ThermaStrain* achieves similar F1 scores (with 1% higher in females than males) and accuracy (almost the same). However, with respect to age and demography, the dataset was limited. Future studies involving

a larger population with diverse ages, sex, and demographic distributions would be highly beneficial. It will allow for a more comprehensive understanding of how the thermal signature of stress manifests across different populations and demographic groups, potentially uncovering any variations or patterns. However, such analysis was out of the scope of this paper.

10 CONCLUSION

Existing studies have examined uni-modal and multimodal thermal stress sensing solutions, each with its advantages and limitations. While uni-modal thermal solutions offer non-intrusive sensing, they may lack effectiveness. On the other hand, multimodal approaches can improve performance but may compromise the non-intrusive nature. *ThermaStrain* combines the benefits of both approaches, providing enhanced stress-sensing performance in a non-intrusive and passive manner. The study collected a comprehensive multimodal thermal stress sensing dataset with diverse stressors and variable distances. Extensive evaluations demonstrated *ThermaStrain*'s ability to generalize and adapt to unknown scenarios, conditions, and environments. These evaluations validated *ThermaStrain*'s fidelity to the co-teaching paradigm and its capacity to enhance stress sensing.

REFERENCES

[1] Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.

[2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[3] Mustafa MM Al Qudah, Ahmad SA Mohamed, and Syaheerah L Lutfi. 2021. Affective State Recognition Using Thermal-Based Imaging: A Survey. *Comput. Syst. Sci. Eng.* 37, 1 (2021), 47–62.

[4] Ane Alberdi, Asier Aztiria, and Adrian Basarab. 2016. Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of biomedical informatics* 59 (2016), 49–75.

[5] Ane Alberdi, Asier Aztiria, Adrian Basarab, and Diane J Cook. 2018. Using smart offices to predict occupational stress. *International Journal of Industrial Ergonomics* 67 (2018), 13–26.

[6] Jerone TA Andrews, Dora Zhao, William Thong, Apostolos Modas, Orestis Papakyriakopoulos, Shruti Nagpal, and Alice Xiang. 2023. Ethical considerations for collecting human-centric image datasets. *arXiv preprint arXiv:2302.03629* (2023).

[7] Apple. 2023. Apple Watch. https://support.apple.com/.

[8] Jeroen HM Bergmann, Vikesh Chandaria, and Alison McGregor. 2012. Wearable and implantable sensors: The patient's perspective. *Sensors* 12, 12 (2012), 16695–16709.

[9] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.

[10] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1 (2013), 1017–1034.

[11] Anne-Marie Brouwer and Maarten A Hogervorst. 2014. A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). *Frontiers in neuroscience* 8 (2014), 224.

[12] Ruud M Buijs and Corbert G Van Eden. 2000. The integration of stress by the hypothalamus, amygdala and prefrontal cortex: balance between the autonomic nervous system and the neuroendocrine system. In *Progress in brain research*. Vol. 126. Elsevier, 117–132.

[13] Sara Campanella, Ayham Altaleb, Alberto Belli, Paola Pierleoni, and Lorenzo Palma. 2023. A Method for Stress Detection Using Empatica E4 Bracelet and Machine-Learning Techniques. *Sensors* 23, 7 (2023), 3565.

[14] Yekta Said Can, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. 2019. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors* 19, 8 (2019), 1849.

[15] Gavin C Cawley and Nicola LC Talbot. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research* 11 (2010), 2079–2107.

[16] CDW. 2023. Future Proofing & New Work Dynamic. Retrieved July, 2023 from https://shorturl.at/ehAGX

[17] Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer Chayes, Levent Sagun, and Riccardo Zecchina. 2019. Entropy-sgd: Biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment* 2019, 12 (2019), 124018.

[18] Sheng-Yang Chiu, Yu-Ting Huang, Chieh-Ting Lin, Yu-Chee Tseng, Jen-Jee Chen, Meng-Hsuan Tu, Bo-Chen Tung, and YuJou Nieh. 2023. Privacy-preserving video conferencing via thermal-generative images. In *2023 IEEE International Conference on Robotics and*

Automation (ICRA). IEEE, 9478–9485.
[19] Youngjun Cho and Nadia Bianchi-Berthouze. 2019. Physiological and affective computing through thermal imaging: A survey. *arXiv preprint arXiv:1908.10307* (2019).
[20] Youngjun Cho, Simon J Julier, and Nadia Bianchi-Berthouze. 2019. Instant stress: detection of perceived mental stress through smartphone photoplethysmography and thermal imaging. *JMIR mental health* 6, 4 (2019), e10140.
[21] Pau Climent-Pérez and Francisco Florez-Revuelta. 2021. Protection of visual privacy in videos acquired with RGB cameras for active and assisted living applications. *Multimedia Tools and Applications* 80, 15 (2021), 23649–23664.
[22] Samantha L Connolly and Lauren B Alloy. 2018. Negative event recall as a vulnerability for depression: Relationship between momentary stress-reactive rumination and memory for daily life stress. *Clinical Psychological Science* 6, 1 (2018), 32–47.
[23] Carl B Cross, Julie A Skipper, and Douglas T Petkie. 2013. Thermal imaging to detect physiological indicators of stress in humans. In *Thermosense: thermal infrared applications XXXV*, Vol. 8705. SPIE, 141–155.
[24] Sally S Dickerson and Margaret E Kemeny. 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin* 130, 3 (2004), 355.
[25] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
[26] Veronika Engert, Arcangelo Merla, Joshua A Grant, Daniela Cardone, Anita Tusche, and Tania Singer. 2014. Exploring the use of thermal infrared imaging in human stress research. *PloS one* 9, 3 (2014), e90782.
[27] Hyukmin Eum, Jeisung Lee, Changyong Yoon, and Mignon Park. 2013. Human action recognition for night vision using temporal templates with infrared thermal camera. In *2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, 617–621.
[28] Victor Foo Siang Fook, Pham Viet Thang, That Mon Htwe, Qiu Qiang, Aung Aung Phyo Wai, Maniyeri Jayachandran, Jit Biswas, and Philip Yap. 2007. Automated recognition of complex agitation behavior of dementia patients using video camera. In *2007 9th International Conference on e-Health Networking, Application and Services*. IEEE, 68–73.
[29] Mathieu Pagé Fortin and Brahim Chaib-Draa. 2019. Multimodal Sentiment Analysis: A Multitask Learning Approach.. In *ICPRAM*. 368–376.
[30] Mihai Gavrilescu and Nicolae Vizireanu. 2019. Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System. *Sensors* 19, 17 (2019), 3693.
[31] Sayandeep Ghosh, Seongki Kim, Muhammad Fazal Ijaz, Pawan Kumar Singh, and Mufti Mahmud. 2022. Classification of mental stress from wearable physiological sensors using image-encoding-based deep neural network. *Biosensors* 12, 12 (2022), 1153.
[32] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P.G. Simos, K. Marias, and M. Tsiknakis. 2017. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control* 31 (2017), 89 – 101. https://doi.org/10.1016/j.bspc.2016.06.020
[33] Ian J Goodfellow, Oriol Vinyals, and Andrew M Saxe. 2014. Qualitatively characterizing neural network optimization problems. *arXiv preprint arXiv:1412.6544* (2014).
[34] William K Goodman, Johanna Janson, and Jutta M Wolf. 2017. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology* 80 (2017), 26–35.
[35] Erin Griffiths, Salah Assana, and Kamin Whitehouse. 2018. Privacy-preserving image processing with binocular thermal cameras. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–25.
[36] James J Gross and Hooria Jazaieri. 2014. Emotion, emotion regulation, and psychopathology: An affective science perspective. *Clinical Psychological Science* 2, 4 (2014), 387–401.
[37] Aline C Gubrium, Amy L Hill, and Sarah Flicker. 2014. A situated practice of ethics for participatory visual and digital methods in public health research and practice: A focus on digital storytelling. *American journal of public health* 104, 9 (2014), 1606–1614.
[38] Constance Hammen. 2005. Stress and depression. *Annu. Rev. Clin. Psychol.* 1 (2005), 293–319.
[39] Margot Hanley, Apoorv Khandelwal, Hadar Averbuch-Elor, Noah Snavely, and Helen Nissenbaum. 2020. An ethical highlighter for people-centric dataset creation. *arXiv preprint arXiv:2011.13583* (2020).
[40] Emily C Helminen, Melissa L Morton, Qiu Wang, and Joshua C Felver. 2019. A meta-analysis of cortisol reactivity to the Trier Social Stress Test in virtual environments. *Psychoneuroendocrinology* 110 (2019), 104437.
[41] Seongsil Heo, Sunyoung Kwon, and Jaekoo Lee. 2021. Stress detection with single PPG sensor by orchestrating multiple denoising and peak-detecting methods. *IEEE Access* 9 (2021), 47777–47785.
[42] Katherine A Herborn, James L Graves, Paul Jerem, Neil P Evans, Ruedi Nager, Dominic J McCafferty, and Dorothy EF McKeegan. 2015. Skin temperature reveals the intensity of acute stress. *Physiology & behavior* 152 (2015), 225–230.
[43] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Flat minima. *Neural computation* 9, 1 (1997), 1–42.
[44] Bonnie D Hodge, Terrence Sanvictores, and Robert T Brodell. 2018. Anatomy, skin sweat glands. (2018).

[45] Daniel Jiwoong Im, Michael Tao, and Kristin Branson. 2016. An empirical analysis of the optimization of deep network loss surfaces. *arXiv preprint arXiv:1612.04010* (2016).

[46] Talha Iqbal, Andrew J Simpkin, Davood Roshan, Nicola Glynn, John Killilea, Jane Walsh, Gerard Molloy, Sandra Ganly, Hannah Ryman, Eileen Coen, et al. 2022. Stress Monitoring Using Wearable Sensors: A Pilot Study and Stress-Predict Dataset. *Sensors* 22, 21 (2022), 8135.

[47] CA James, AJ Richardson, PW Watt, and NS Maxwell. 2014. Reliability and validity of skin temperature measurement by telemetry thermistors and a thermal camera during exercise in the heat. *Journal of thermal biology* 45 (2014), 141–149.

[48] Hayeon Jeong, Heepyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch wearing behavior analysis: a longitudinal study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–31.

[49] Marcin Jukiewicz, Paweł Łupkowski, Radomir Majchrowski, Joanna Marcinkowska, and Dawid Ratajczyk. 2021. Electrodermal and thermal measurement of users' emotional reaction for a visual stimuli. *Case Studies in Thermal Engineering* 27 (2021), 101303.

[50] Kenji Kawaguchi, Leslie Pack Kaelbling, and Yoshua Bengio. 2017. Generalization in deep learning. *arXiv preprint arXiv:1710.05468* (2017).

[51] N Keshan, PV Parimi, and Isabelle Bichindaritz. 2015. Machine learning for stress detection from ECG signals in automobile drivers. In *2015 IEEE International conference on big data (Big Data)*. IEEE, 2661–2669.

[52] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. 2016. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836* (2016).

[53] Reza Khosrowabadi. 2018. Stress and Perception of Emotional Stimuli: Long-term Stress Rewiring the Brain. *Basic and clinical neuroscience* 9, 2 (2018), 107.

[54] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235.

[55] JongBae Kim. 2019. Pedestrian detection and distance estimation using thermal camera in night time. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 463–466.

[56] Yoonkyoung Kim, Yosep Park, Jinman Kim, and Eui Chul Lee. 2018. Remote heart rate monitoring method using infrared thermal camera. *Int. J. Eng. Res. Technol* 11, 3 (2018), 493–500.

[57] Ayca Kirimtat, Ondrej Krejcar, Ali Selamat, and Enrique Herrera-Viedma. 2020. FLIR vs SEEK thermal cameras in biomedicine: comparative diagnosis through infrared thermography. *BMC bioinformatics* 21, 2 (2020), 1–10.

[58] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. 1993. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.

[59] Alexandra König, Carlos Fernando Crispim Junior, Alexandre Derreumaux, Gregory Bensadoun, Pierre-David Petit, François Bremond, Renaud David, Frans Verhey, Pauline Aalten, and Philippe Robert. 2015. Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients. *Journal of Alzheimer's Disease* 44, 2 (2015), 675–685.

[60] Ethan Kross, David Gard, Patricia Deldin, Jessica Clifton, and Ozlem Ayduk. 2012. "Asking why" from a distance: Its cognitive and emotional consequences for people with major depressive disorder. *Journal of abnormal psychology* 121, 3 (2012), 559.

[61] Alan T Krzywicki, Gary G Berntson, and Barbara L O'Kane. 2014. A non-contact technique for measuring eccrine sweat gland activity using passive thermal imaging. *International journal of psychophysiology* 94, 1 (2014), 25–34.

[62] Satish Kumar, ASM Iftekhar, Michael Goebel, Tom Bullock, Mary H MacLean, Michael B Miller, Tyler Santander, Barry Giesbrecht, Scott T Grafton, and BS Manjunath. 2021. StressNet: detecting stress in thermal videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 999–1009.

[63] Angeliki Kylili, Paris A Fokaides, Petros Christou, and Soteris A Kalogirou. 2014. Infrared thermography (IRT) applications for building diagnostics: A review. *Applied Energy* 134 (2014), 531–549.

[64] Juwon Lee, Megan Lam, and Caleb Chiu. 2019. Clara: design of a new system for passive sensing of depression, stress and anxiety in the workplace. In *Pervasive Computing Paradigms for Mental Health: 9th International Conference, MindCare 2019, Buenos Aires, Argentina, April 23–24, 2019, Proceedings 9*. Springer, 12–28.

[65] Kwangyoung Lee, Hyewon Cho, Kobiljon Toshnazarov, Nematjon Narziev, So Young Rhim, Kyungsik Han, YoungTae Noh, and Hwajung Hong. 2020. Toward future-centric personal informatics: Expecting stressful events and preparing personalized interventions in stress management. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[66] Alessandro Leone, Gabriele Rescio, Pietro Siciliano, Alessandra Papetti, Agnese Brunzini, and Michele Germani. 2020. Multi sensors platform for stress monitoring of workers in smart manufacturing context. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 1–5.

[67] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. 2018. Visualizing the loss landscape of neural nets. *Advances in neural information processing systems* 31 (2018).

[68] Yi Li, Rameswar Panda, Yoon Kim, Chun-Fu Richard Chen, Rogerio S Feris, David Cox, and Nuno Vasconcelos. 2022. VALHALLA: Visual Hallucination for Machine Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5216–5226.

[69] Kevin Lin, Lijuan Wang, Kun Luo, Yinpeng Chen, Zicheng Liu, and Ming-Ting Sun. 2020. Cross-domain complementary learning using pose for multi-person part segmentation. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 3 (2020), 1066–1078.

[70] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[71] Bing Liu. 2017. Many facets of sentiment analysis. In *A practical guide to sentiment analysis*. Springer, 11–39.

[72] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*. 4765–4774.

[73] Dominique Makowski, Tam Pham, Zen J Lau, Jan C Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods* (2021), 1–8.

[74] Peter Mantello and Manh-Tung Ho. 2023. Emotional AI and the future of wellbeing in the post-pandemic workplace. *AI & society* (2023), 1–7.

[75] marcellodebernardi. 2019. loss-landscapes. https://github.com/marcellodebernardi/loss-landscapes.

[76] Maurizio Mauri, Valentina Magagnin, Pietro Cipresso, Luca Mainardi, Emery N Brown, Sergio Cerutti, Marco Villamira, and Riccardo Barbieri. 2010. Psychophysiological signals associated with affective states. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 3563–3566.

[77] Pauline Maurice, Ludivine Allienne, Adrien Malaisé, and Serena Ivaldi. 2018. Ethical and social considerations for the introduction of human-centered technologies at work. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. IEEE, 131–138.

[78] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[79] Luca Menghini, Evelyn Gianfranchi, Nicola Cellini, Elisabetta Patron, Mariaelena Tagliabue, and Michela Sarlo. 2019. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology* 56, 11 (2019), e13441.

[80] Arcangelo Merla and Gian Luca Romani. 2007. Thermal signatures of emotional arousal: a functional infrared imaging study. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 247–249.

[81] Shekhar Neema, DM Tripathy, Sweta Mukherjee, Anwita Sinha, Senkadhir Vendhan, and Biju Vasudevan. 2021. Infrared thermography in the diagnosis of palmar hyperhidrosis: A diagnostic study. *Medical Journal Armed Forces India* (2021).

[82] Thu Nguyen, Khang Tran, and Hung Nguyen. 2018. Towards thermal region of interest for human emotion estimation. In *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 152–157.

[83] Søren Z Nielsen, Rikke Gade, Thomas B Moeslund, and Hans Skov-Petersen. 2014. Taking the temperature of pedestrian movement in public spaces. *Transportation Research Procedia* 2 (2014), 660–668.

[84] Simon Ollander, Christelle Godin, Aurélie Campagne, and Sylvie Charbonnier. 2016. A comparison of wearable and stationary sensors for stress detection. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 004362–004366.

[85] Ioannis Pavlidis, Norman L Eberhardt, and James A Levine. 2002. Seeing through the face of deception. *Nature* 415, 6867 (2002), 35–35.

[86] Ioannis Pavlidis, Panagiotis Tsiamyrtzis, Dvijesh Shastri, Avinash Wesley, Yan Zhou, Peggy Lindner, Pradeep Buddharaju, Rohan Joseph, Anitha Mandapati, Brian Dunkin, et al. 2012. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports* 2, 1 (2012), 305.

[87] Verónica Pérez-Rosas, Alexis Narvaez, Mihai Burzo, and Rada Mihalcea. 2013. Thermal imaging for affect detection. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*. 1–4.

[88] David Perpetuini, Damiano Formenti, Daniela Cardone, Chiara Filippini, and Arcangelo Merla. 2021. Regions of interest selection and thermal imaging data analysis in sports and exercise science: a narrative review. *Physiological Measurement* 42, 8 (2021), 08TR01.

[89] Rosalind W Picard. 2016. Automating the recognition of stress and emotion: From lab to real-world impact. *IEEE MultiMedia* 23, 3 (2016), 3–7.

[90] Colin Puri, Leslie Olson, Ioannis Pavlidis, James Levine, and Justin Starren. 2005. StressCam: Non-contact measurement of users' emotional states through thermal imaging. *Proceedings of the 2005 ACM Conference on Human Factors in Computing Systems* 2, 1725–1728. https://doi.org/10.1145/1056808.1057007

[91] Siyuan Qiao, Liang-Chieh Chen, and Alan Yuille. 2020. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution. *arXiv preprint arXiv:2006.02334* (2020).

[92] Jose Ignacio Priego Quesada, Natividad Martínez Guillamón, Rosa Ma Cibrián Ortiz de Anda, Agnes Psikuta, Simon Annaheim, René Michel Rossi, José Miguel Corberán Salvador, Pedro Pérez-Soriano, and Rosario Salvador Palmer. 2015. Effect of perspiration on skin temperature measurements by infrared thermography and contact thermometry during aerobic cycling. *Infrared Physics & Technology* 72 (2015), 68–76.

[93] Vandana Rajan, Alessio Brutti, and Andrea Cavallaro. 2021. Robust Latent Representations Via Cross-Modal Translation and Alignment. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4315–4319.

[94] Edna Maria Vissoci Reiche, Sandra Odebrecht Vargas Nunes, and Helena Kaminami Morimoto. 2004. Stress, depression, the immune system, and cancer. *The lancet oncology* 5, 10 (2004), 617–625.

[95] Adrian Rosebrock. 2016. Intersection over Union (IoU) for object detection. *Diambil kembali dari PYImageSearch: https://www. pyimagesearch. com/2016/11/07/intersection-over-union-iou-for-object-detection* (2016).

[96] Nazreen Rusli, Shahrul Naim Sidek, Hazlina Md Yusof, Nor Izzati Ishak, Madihah Khalid, and Ahmad Aidil Arafat Dzulkarnain. 2020. Implementation of wavelet analysis on thermal images for affective states recognition of children with autism spectrum disorder. *IEEE Access* 8 (2020), 120818–120834.

[97] SafelyYou. 2022. Transform care delivery with world-leading AI + clinical expertise. Retrieved July, 2023 from https://www.safely-you.com/

[98] Asif Salekin, Hongning Wang, Kristine Williams, and John Stankovic. 2017. Dave: detecting agitated vocal events. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 157–166.

[99] Zhanna Sarsenbayeva, Niels van Berkel, Danula Hettiachchi, Weiwei Jiang, Tilman Dingler, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2019. Measuring the effects of stress on mobile interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–18.

[100] Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. 2010. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion* 24, 7 (2010), 1153–1172.

[101] ProTech Security. 2023. How Thermal Cameras for Businesses Can Keep Employees and Customers Safe. Retrieved July, 2023 from https://protechsecurity.com/how-thermal-cameras-for-businesses-can-keep-employees-and-customers-safe/

[102] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2009. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine* 14, 2 (2009), 410–417.

[103] Lloyd S Shapley. 1953. Stochastic games. *Proceedings of the national academy of sciences* 39, 10 (1953), 1095–1100.

[104] Harshit Sharma, Yi Xiao, Victoria Tumanova, and Asif Salekin. 2022. Psychophysiological Arousal in Young Children Who Stutter: An Interpretable AI Approach. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–32.

[105] Ben Shneiderman. 2020. Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–31.

[106] Saurabh Sonkusare, David Ahmedt-Aristizabal, Matthew J Aburn, Vinh Thai Nguyen, Tianji Pang, Sascha Frydman, Simon Denman, Clinton Fookes, Michael Breakspear, and Christine C Guo. 2019. Detecting changes in facial temperature induced by a sudden auditory stimulus based on deep learning-assisted face tracking. *Scientific reports* 9, 1 (2019), 4729.

[107] Nattapong Thammasan, Koichi Moriyama, Ken-ichi Fukui, and Masayuki Numao. 2017. Familiarity effects in EEG-based emotion recognition. *Brain informatics* 4 (2017), 39–50.

[108] Victoria Tumanova and Nicole Backes. 2019. Autonomic nervous system response to speech production in stuttering and normally fluent preschool-age children. *Journal of Speech, Language, and Hearing Research* 62, 11 (2019), 4030–4044.

[109] Marieke van Dooren, Joris H Janssen, et al. 2012. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & behavior* 106, 2 (2012), 298–304.

[110] Maarten Vandersteegen. 2018. SEEK thermal compact camera driver supporting the thermal Compact, thermal CompactXR and and thermal CompactPRO. https://github.com/maartenvds/libseek-thermal

[111] Wilhelm Von Rosenberg, Theerasak Chanwimalueang, Tricia Adjei, Usman Jaffer, Valentin Goverdovsky, and Danilo P Mandic. 2017. Resolving ambiguities in the LF/HF ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. *Frontiers in physiology* 8 (2017), 360.

[112] Rahee Walambe, Pranav Nayak, Ashmit Bhardwaj, and Ketan Kotecha. 2021. Employing multimodal machine learning for stress detection. *Journal of Healthcare Engineering* 2021 (2021), 1–12.

[113] Qianqian Wang, Zhiqiang Tao, Wei Xia, Quanxue Gao, Xiaochun Cao, and Licheng Jiao. 2022. Adversarial multiview clustering networks with adaptive fusion. *IEEE transactions on neural networks and learning systems* (2022).

[114] Zhikang Wang, Feng Zhu, Shixiang Tang, Rui Zhao, Lihuo He, and Jiangning Song. 2022. Feature erasing and diffusion network for occluded person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4754–4763.

[115] Kay Wright and Swaran Singh. 2022. Reducing falls in dementia inpatients using vision-based technology. *Journal of Patient Safety* 18, 3 (2022), 177.

[116] Jiacheng Yang. 2022. *Enabling Privacy-Preserving Model Personalization via On-Device Incremental Training*. Ph. D. Dissertation. University of Toronto (Canada).

[117] Bin Yu, Mathias Funk, Jun Hu, Qi Wang, and Loe Feijs. 2018. Biofeedback for everyday stress management: A systematic review. *Frontiers in ICT* 5 (2018), 23.

[118] Huijun Zhang, Ling Feng, Ningyun Li, Zhanyu Jin, and Lei Cao. 2020. Video-Based Stress Detection through Deep Learning. *Sensors* 20, 19 (2020), 5552.

[119] Jing Zhang, Hang Yin, Jiayu Zhang, Gang Yang, Jing Qin, and Ling He. 2022. Real-time mental stress detection using multimodality expressions with a deep learning framework. *Frontiers in Neuroscience* 16 (2022).

[120] Zhuo Zheng, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. 2021. Deep multisensor learning for missing-modality all-weather mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* 174 (2021), 254–264.

[121] Kaiyang Zhou and Tao Xiang. 2019. Torchreid: A library for deep learning person re-identification in pytorch. *arXiv preprint arXiv:1910.10093* (2019).

[122] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. 2019. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*. 3702–3712.

[123] Lili Zhu, Pai Chet Ng, Yuanhao Yu, Yang Wang, Petros Spachos, Dimitrios Hatzinakos, and Konstantinos N Plataniotis. 2022. Feasibility study of stress detection with machine learning through eda from wearable devices. In *ICC 2022-IEEE International Conference on Communications*. IEEE, 4800–4805.

[124] The Guardian Zoë Corbyn. 2021. The future of elder care is here – and it's artificial intelligence. Retrieved July, 2023 from https://www.theguardian.com/us-news/2021/jun/03/elder-care-artificial-intelligence-software

A APPENDIX

A.1 Loss Landscape visualization

Li et al. [67] showed that visualizing the loss landscape for a neural network provides a richer understanding of how the different model architecture and other design choices influence the optimization of the loss function. While generating the loss landscapes, we intend to visualize the impact of model parameters or θ, which is a high dimensional quantity. For interpretability, it is required to reduce its dimensionality to a one or two-dimensional hyperspace. To address this challenge prior works[33, 45, 67] choose a starting point in the parameter subspace θ and choose two random Gaussian directions vectors given by δ and η and plot the graph for:

$$f(\alpha, \beta) = L(\theta + \alpha\delta + \beta\eta) \tag{8}$$

This equation generates 3D visualization of the loss landscape with XY region bounded by two scalar quantities or step sizes α (x-axis) and β (y-axis) and corresponding loss for the $L(\theta + \alpha\delta + \beta\eta)$ as the z-axis. Furthermore, Li et al. [67] suggest using filter normalized direction vectors δ and η helps to capture the natural distance scale of the loss surfaces (details can be found in the original work[67]). We used the *loss-landscapes* library [75] which also uses filter-normalization appraoch [67] to generate the 3D loss landscape plots for the best thermal baseline, and our ThermaStrain approach which is shown in the paper (Figure 7). We used the cross-entropy loss for the graph generation, and both the graphs were generated for a randomly selected participant from the validation set for step sizes ($\alpha = 40, \beta = 40$).