# What not to do in Facial Infrared Thermographic Measurements: A Post Data Enhancement

Pike, Matthewa., Yeboah, Siegfriedb,[[1]](#footnote-1)\*., and Fu Xianyunb

1. *School of Computer Science, University of Nottingham Ningbo China, Ningbo 315100, P.R. China. Email:* [*Matthew.Pike@nottingham.edu.cn*](mailto:Matthew.Pike@nottingham.edu.cn)*. ORCID:* [*https://orcid.org/0000-0002-1543-0148*](https://orcid.org/0000-0002-1543-0148)
2. *Department of Architecture and Built Environment, The University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo 315100, P.R. China. Email:* [*fuxianyun0217@gmail.com*](mailto:fuxianyun0217@gmail.com) *ORCID:* [*https://orcid.org/0000-0002-4287-8693*](https://orcid.org/0000-0002-4287-8693)

## Abstract

The accuracy of infrared thermographic measurements depends on several factors, including movement of target. In this study, accuracy of nose tip temperatures obtained in a mental workload assessment using a thermal imaging camera were impacted by participants’ movement and camera zooming/panning. To correct these measurement errors, we compared manual facial landmark identification techniques using data labelling software with an automated deep learning-based approach utilised for facial landmark tracking and evaluated both against the built-in tracking features of the thermal camera, Thermal Spot Tracking. Using the Manual Thermal Landmark Annotation measurements as the ground truth, our results show that the Automated Facial Feature Tracking approach, which is the AI based approach performed better than the Thermal Spot Tracking as it matched comparatively more spatial coordinates and temperature datapoints as well as showed comparatively lower mean relative error. The study highlights the potential of AI in enhancing the accuracy of thermographic measurements, particularly in applications involving facial temperature analysis.

**Keyword**s: *Deep Learning; Convolutional Neural Network; Facial Landmark Detection; Infrared Thermography.*

**Abbreviations [[2]](#footnote-2)**

## 1. Introduction

Mental Workload (MWL) is one characterisation of cognitive work that is often used to model an operator’s cognitive engagement in a task. It is a theoretical model suggested to share a strong relationship with human performance [1], [2]. MWL can be indicative of an operator who is disengaged/underworked (low MWL), engaged (high-MWL), overloaded (excessively high-MWL) or fatigued (gradual/continual decline in MWL). Existing research utilise non-invasive, but intrusive on-body sensors, such as functional Near Infrared Spectroscopy (fNIRS) [2], [3], Electroencephalogram (EEG) [4] and Heart-Rate Variability (HRV) [5] to model cognitive work, often characterised through the application of a theoretical MWL model [6]. On-body sensors are not without limitations; however, a variety of factors may influence a sensor’s reliability in real world environments. Issues may include participant discomfort, sensor contact quality, motion derived artefacts, unreliable connectivity, and an observation bias [2]. Non-intrusive approaches to predicting operator MWL show promise to broader adoption of MWL in realistic context. Facial Thermography (FT) is one approach to measuring MWL in a non-intrusive manner and has several advantages over existing on-body sensors for predicting MWL. FT is a non-intrusive, remote-sensing measure, meaning it does not impair nor disrupt the participant. Further, FT can operate in dark or poorly lit environments and is a relatively affordable measure to deploy. Existing work has shown FT to be sensitive to changes in MWL [7], [8], which has been attributed to the vasoconstriction response of the autonomic nervous system [9].

Infrared thermography is a non-invasive approach to measure surface temperatures capable of producing measurements with high spatial and temporal resolution [10]. It is used to measure and convert mid to long-wave infrared radiation from objects to temperature, presented via rendered thermal map in false colour as digitized images or high-speed videos [11] . Playà-Montmany and Tattersall [12] found infrared thermography to be increasingly used as an analytical approach in understanding how plants and animals interact with their thermal environment, although there is the possibility of yielding inaccurate results due to deviations from the ‘correct’ value. Tian et al. [13] analysed the influencing factors of an infrared camera in facial skin temperature measurements and found that the main influencing factors were measurement differences in the skin temperature, the surrounding air temperature, and the position of the camera. Common sources of error in measurements using infrared thermography include the use of default emissivity settings leading to under estimation in cool or warm objects; inaccuracy in the reflected background environmental temperature; air moisture content; thermal image view angle and movement; field of view of lens; and measurement distance [10], [11], [14], [15]. Jiao et al [15] found that without compensation treatment, the temperature image measured in the angle of view of 74° to 76° may introduce additional error. Using an infrared thermal imager to facilitate the screening of animals showed a difference of more than 2°C compared with that measured in the angle of view of 0°. This, they found, was due to random motion and small surface temperature variation of the animals, making the influence of the angle of view on temperature measurement significant. Surface temperature estimation errors have also been found to occur due to the difficulty in maintaining fixed angles on moving targets and because biological surfaces are rarely perfectly flat geometric shapes [12]. Playà-Montmany and Tattersall [12] found that changing angles in thermography alters the estimated surface temperature if a constant emissivity is assumed, as images they captured of an Atlantic puffin, 0.97s apart, showed that a steep change in angle due to head movement altered the bill temperature by ~1°C. Bell et al. [14] investigated the use of thermal imaging as a technique to assess core body temperature in pre-weaned artificially reared calves. They found that the movement of the calf's head, during measurements, led to variation in the angle at which the thermal image was taken, resulting in some variability in the readings obtained. Li et al. [16] found that a significant limit to the applicability of thermal cameras in the real operational settings is the fact that subjects can move around at will, leading to errors in the data collected. Fajkus et al. [17] developed a non-invasive hybrid multichannel fiber-optic sensor system for basic vital sign monitoring such as body temperature, respiratory rate, and heart rate. The study found that measurement sensitivity was affected when the monitored person for example changes his/her position (minor artefacts), performs additional movements, coughs, etc. (major artefacts). Metzmacher et al. [18] carried out real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment. They found that a measurement point with a unit radius can lead to a high amount of noise since the measurement is subject to inaccuracies in tracking, subtle movements, and measurement noise from the thermal camera's sensor itself. They also found that crossing segment boundaries when measuring a desired body segment may lead to higher measurement error.

Measurement of temperature using thermography involves the scanning across a body or scene, and thus provides information about temperature distributions instead of a single point at the focal point of the instrument [19]. Thermal imaging sensor measures the entire environment within the field of view, which means that while the image may be focused on a particular region, the entire environment still impinges upon and/or influences the radiation which the sensor detects [11]. Schweikert et al. [10] observed that the accurate determination of the surface temperature of a moving object is, however, of great interest to a range of practical applications. Ranjan & Scott [20] used thermographic imaging to dynamically detect and predict thermal comfort and found that the tracking capabilities of the thermographic camera will be crucial in capturing the face of a person when they are at their desk in an office. Tattersall [11] found that although thermal imaging cameras can capture images at any angle of incidence perpendicular to the object, they sometimes capture images at very low angle of incidence influencing the recorded temperature.

We surmise from the various studies that camera positioning needs careful consideration to obtain near accurate thermographic measurements. With the challenges outlined, effective use of thermal imaging cameras may potentially be limited to cases of stationary/observational work types where movements are not significant to disrupt measurements. Artificial Intelligence (AI) has been extensively applied in industrial contexts using thermographic imagery, especially for fault and anomaly detection. Akram et al., [21] developed a Deep Learning (DL) model for detecting lab-induced defects in Photovoltaic modules monitored using a thermographic imager with a classification accuracy of 98.67%. The model was further fine-tuned through transfer learning, resulting in improved accuracy of 99.23%. This model can run on common computing hardware and has real-time prediction performance. D'Orazio et al., [22] analysed composite materials Nonex and Syncore within an aircraft using a ML based neural network to recognise different defects such as water intrusion and holes caused by impact. Yang et al, [23] presented a Convolutional-Neural Network (CNN), an ML approach, to identify cracks in heated steel surfaces by utilising temperature differences as indicators of varying crack depths. The work produced a dataset of 3000 labelled images which were used to train the CNN with an accuracy of 95.54%.

Beyond industrial applications, there has been a growing interest in AI and thermographic imaging for health-centered research. Mambou et al., [24] presented an innovative Machine Learning (ML) approach that uses Support Vector Machines (SVMs) to analyse frontal thermogram images and differentiate between healthy and cancerous breast tissue. Similarly, AI and thermographic imaging have found several applications in analysing respiratory behaviours, such as stress recognition [25], air leakage detection [26], and real-time abnormal respiratory recognition [27]. The combination of thermographic imaging and AI is particularly appealing in the medical field due to its non-invasive and non-intrusive diagnostic capabilities.

In this study, we focus on integrating thermographic imaging with AI to enable more flexible and realistic use-case scenarios. Contrasting with most studies discussed here, which primarily involve stationary objects like airplane structures and immobile industrial machinery, our approach seeks to overcome these limitations. This is particularly evident in medical settings, where our approach may present an approach towards less restrictive conditions, unlike traditional practices where patients must remain completely still or maintain unnatural poses during imaging procedures. Existing works have made progress towards this. Akula et al., [28] demonstrated the use of thermal imaging and CNN to perform human activity recognition in ambient assisting living environments, typically associated with assisting the elderly and disabled. The proposed system was able to classify six actions (sitting, falling, etc) from thermal imagery with 87.44% accuracy. Bhattacharyya et al., [29] presented a DL model called IRFacExNet (InfraRed Facial Expression Network) for performing facial expression recognition from infrared images which can discriminate between 8 classes of expression (Happy, Neutral, Sad, etc) with 88.43% recognition accuracy. Lin et al.,[30] introduced a Random Forest approach for facial recognition using thermal imagery, effectively handling a variety of conditions including the presence of objects like glasses and face masks during the experiments. Kuzdeuov et al.,[31] provide an extensive dataset of labelled facial features from 147 subject in both controlled and uncontrolled environments. This dataset, along with the model developed from it, served as the foundational basis for the model employed in our current study.

In this present study, we employed an artificial intelligence (AI) approach, known as Automated Facial Feature Tracking (AFFT), to improve thermographic temperature measurements. To demonstrate this, we used a case study involving thermographic measurements to obtain facial temperature distribution to assess participant mental workload (MWL) whilst performing a memory-based task. We then proceed to show how, using a pretrained neural network, we were able to enhance the thermographic images post measurements. We then compared the performance of the AFFT approach with the inherent tracking capabilities of the thermal camera and a manual approach using image annotation software. The paper is structured as follows: Section 2 presents the case study where the challenges in the thermographic measurements are outlined. In Section 3, our materials and methods section outline our post data enhancement strategy, measurement processes and data analysis approach. Section 4 then presents the results and analysis relating to the performance of the TST and AFFT approaches compared to our ground truth measurement approach, MTLA. Here, we also present the uncertainty in the measurements, valuable learnings from the study as well as the study limitations and opportunities for future research before we outline our key findings in Section 5, our conclusion of the paper.

## 2. The Case Study: Measurement of Mental Workload (MWL) using Facial Infrared Thermography

### 2.1. Case Study Brief

The study was carried out to explore the relationship between participant's nose tip temperature change and MWL in a realistic (uncontrolled) working environment. It involved investigating an individual’s MWL in realistic working environments using thermographic measurements. 20 participants (10 male, 10 female) were recruited to take part in the study. All participants were between the ages of 19-22 years and were undergraduate students. All participants had normal or corrected vision and reported no history of head trauma or brain injury.

After ethics approval from the University’s faculty ethics committee and participants completing a consent form (See Appendix 1), they were asked to undertake a Dual N-Back test, a task known to vary workload levels successfully [32]. By controlling the number of items in the Dual N-Back test, participants’ nose tip temperatures were expected to change per their cognitive engagement with the task. Generally, it is expected that nose tip temperature decreases with increased MWL [7], [8]. The study was conducted in April 2021 (Spring season), by an undergraduate student (the “student researcher”) collecting the data, under the supervision of faculty academics. Participants provided informed consent and were not compensated for their participation.

### 2.2. Case Study Material and Methods

Each participant completed the Dual N-Back task on a laptop placed on a desk of height 0.75m (See Figures 1 and 2). The office-like environment the task was undertaken had a standard air-conditioner set to room temperature of 24°C. An infrared camera placed 1.4m away from the desk was used to record facial skin temperature at 15-second intervals while participants performed the Dual N-Back task as depicted in Figure 1.

图示

描述已自动生成

*Figure 1 Physical Model Showing Facial Thermography* *Measurement Process*

房间里的桌子上放着电脑

中度可信度描述已自动生成

*Figure 2 Picture of Measurement Location with Setup of Thermal Cameras and Laptop.*

The measurements took place during spring season from 14th to 17th April 2021. Sensors and dataloggers were installed to monitor the indoor environmental parameters such as humidity, stratified air temperatures and air velocity (See Figure 3). Just as MWL, specific facial points are crucial for individual’s thermal comfort assessment. This is because, when the indoor air temperature changes, physiological performance may also change [33], [34]. For instance, Li et al. [16] found that the ears, nose, and cheeks can best reflect the thermal comfort level, with an average accuracy of 85%. Ghahramani et al. [35] measured facial skin temperature (cheekbone, ear, front face, and nose) and found that the room temperature has a high correlation with the average measured facial points.

|  |
| --- |
| 图示  描述已自动生成 |
| *Figure 3 Setup for Indoor Environmental Condition Monitoring* |

The facial skin temperature of participants was measured using FLIR T640 thermal imaging camera. The calibration of the FLIR T640 camera (primary camera for the study) was done by the manufacturer TELEDYNE FLIR hence to ensure measurements were accurate and consistent, a secondary camera, FLIR E75 also calibrated by the manufacturer was used to corroborate the measurements. See Table 1 for the accuracies of both the primary and secondary cameras. Four spots (Spot 1,2 ,3 and 4) on the face, specifically on the forehead, ocular region and nose were chosen for the measurement of the facial temperature distribution (See Figure 4).

|  |
| --- |
|  |
| *Figure 4 Avatar* [36] *with Facial Spots next to Facial Temperature Measurement* |

### 2.3. Case Study Challenges with Facial Temperature Measurements

The model image in Figure 4 represents the ideal results expected from the facial temperature measurements where specific spots were chosen. However, due to movements by subjects, the spots selected to measure facial temperature to determine MWL shifted during measurements. Just as found in studies by Bell et al. [14], Li et al. [16] and Fajkus et al. [17] the movement by participants led to erroneous outcomes during analysis by the student researcher. This situation was further worsened by the manual annotation of the spots on the thermographic images representing the nose tip temperature. In Figure 5, some of the measurements that led to the erroneous results due to movements are shown.

|  |  |
| --- | --- |
| 图示  描述已自动生成 | 图形用户界面  描述已自动生成 |
| 图形用户界面  中度可信度描述已自动生成 | 图形用户界面  低可信度描述已自动生成 |
| 图形用户界面  描述已自动生成 | 图形用户界面  中度可信度描述已自动生成 |
| 图示  低可信度描述已自动生成 | 图片包含 图示  描述已自动生成 |
| 图片包含 图示  描述已自动生成 | 图示  描述已自动生成 |
| *Figure 5 Erroneous Facial Temperature Measurement Points* | |

## 3. Material and Methods

### 3.1. Post Data Enhancement

The purpose of the post data enhancement was to minimize the impact of human errors in the data by reducing system detection errors due to participants’ movements which changed static spot locations during measurements. In the post data enhancement of the results, the facial temperature data was obtained manually and set as the ground truth, against which the performance of automatic annotation approaches were compared to. This was possible because infrared cameras capture the entire environment within the field of view and the sensor detects the radiation from the entire environment [11]. Hence it was possible to retrieve specific temperature data from any surface captured by the camera. We employed two automatic methods in our study: Static Landmark Detection, also known as TST, and AFFT, to generate data for comparison with the ground truth, MTLA dataset.

The static landmark detection approach utilised the thermographic camera’s functionality to record “spots” during capture. In this post data enhancement only spot 4 from the four spots (Spot 1, 2 ,3 and 4 shown in Figure 4) depicting the nose tip temperature in the case study work carried out by the student researcher was used in the analysis. It is worthy to note that the spot detection capabilities of the FLIR T640 thermal imaging camera does not dynamically track the subject subsequently impacting the accuracy of the facial temperature measurements when participants moved, as can be seen in the images in Figure 5. Given the nature of the task, participants were required to sit stationary throughout the study, focussed on a single area of the laptop screen, limiting head movement. Static spots serve as a quick and simple way of implementing this monitoring and may have value in particular use cases. However, the inherent limitation of the FLIR T640 thermal imaging camera to track and maintain static spot location during measurement resulted in different facial point temperatures being recorded for designated static points. This obviously influenced participants’ physiological condition as being aware that they were to remain almost still for the duration of the measurement potentially influenced the results obtained for their MWL. In more realistic scenarios, however, it is unlikely that participants will remain perfectly still for extended periods of time. To this end, automated methods for tracking become appealing functionality for applying this approach in realistic working environments.

Extracting facial landmarks from imagery is one typical application of machine learning technologies for colour image use cases. Zhang et al. [37], demonstrated the application of deep multi-task learning approaches to the remote identification of facial landmarks for images with significant occlusion and facial pose variations. A broad survey of related literature compiled by Wu and Ji. [38] shows that fewer approaches, have been applied to thermographic imagery.

### 3.2. Measurement Process

The first stage of each facial thermographic data collection process involved the researcher, volunteering participant, thermal camera, and software as shown in Figure 6. The measurement process was in a loop cycle where the researcher instructed each volunteering participant to complete a consent form (See Appendix 1) and then perform the Dual N-Back task while the FLIR T640 thermal imaging camera recorded their facial temperature during the process. The FLIR T640 thermal imaging camera’s spot detection was engaged to track four facial points including the nose tip of each participant. The indoor environmental conditions were monitored to ensure the parameters were within the thermal comfort zone for the participants’, so the environment does not become a factor that negatively influences the integrity of the results obtained. As shown in Figure 6, the second stage of the data collection process involved manual annotations of the thermographic images, establishment of a MTLA dataset, extraction of spot data before using an AFFT to predict nose tip coordinates from each thermographic image.

A screenshot of a diagram

Description automatically generated  
*Figure 6 Sequence diagram for the study and data generation.*

### 3.3. Manual Thermal Landmark Annotation (MTLA)

The ground truth, referred to herein as MTLA, facial temperature dataset was generated via manual annotation of the original thermal images collected during the study. Annotation was performed using the data labelling software RectLabel[[3]](#footnote-3) as shown in Figures 7a and 7b. For each image, the centre crosshair of the top-left corner of the annotating rectangle was placed on the centre point of the participant's nose tip. A representative temperature value was then derived by averaging the values from 9 neighbouring pixels around the crosshair location. Once completed, the annotation data (X, Y pixel coordinates) for each image were exported to a standard file format, XML, for further processing. Annotation correctness was confirmed via visual inspection, performed by a different independent researcher to the one who performed the manual annotation.

Python was then used to read both the generated XML annotations and the original thermal images captured during the study. For each image, the associated annotation data was identified and read from the XML file. Temperature data, for each pixel was extracted using *flirextractor* [[4]](#footnote-4) [39], with temperature values obtained by reading the value specified by the X, Y coordinates stored in the XML file. This data was then stored in an Excel spreadsheet. This procedure leveraged the Python library, Pandas [40], to perform the input/output operations described previously.

|  |  |
| --- | --- |
|  |  |
| *Figure 7a MTLA Annotation using RectLabel1* | *Figure 7b Detailed Image of MTLA Annotation* |

### 3.4. Thermal Spot Tracking (TST)

For this approach, a total of 4 spots were pre-defined for each participant using the spot interface on the FLIR T640 thermal imaging camera as shown in Figure 8. In the case study, the student researcher utilized the camera’s digital zoom function to align the spots accurately at the beginning of the study. For the purposes of this analysis, we are only concerned with spot 4, which was aligned with the tip of the participant’s nose.

|  |
| --- |
|  |
| *Figure 8 Pre-defined Spots on Participants Face for TST approach.* |

Spot coordinates were extracted using *exiftool*[[5]](#footnote-5), an image metadata parsing and writing library. Whilst *exiftool* can extract spot coordinates, it is unable to read temperature measurements associated with a given spot hence the temperature from each image was manually recorded. We note, however, that this process could be automated by combining the X, Y coordinates obtained via *exiftool*, with the temperature values using *flirextractor*, as described above. However, this approach was not used in this study, because of complications introduced from using the digital zoom and panning functionality. The X, Y coordinates obtained from *exiftools* respected the level of digital zoom used by the operator, at the time of capture. However, *flirextractor,* does not respect the level of zoom, and instead returns a temperature value matrix for **all** pixels present in the original image. This resolution difference led to inaccuracies in calculating the actual temperature value at a given point. We were unable to resolve these differences through postprocessing because the amount of manual panning performed by the student researcher was not recorded in the resulting image metadata. Therefore, we manually recorded the temperatures reported in the FLIR images into a spreadsheet.

### 3.5. Automated Facial Feature Tracking (AFFT)

Automatic facial feature tracking was conducted using a DL-based approach. The method employed the YOLOv5m6-Face architecture, which is derived from the YOLO5Face detection model [41], a specialized model for detecting facial features. YOLO5Face itself is based on the general-purpose object detection model YOLOv5 [42]. This model utilizes a CNN to perform object detection and localization, adhering to the principles of the YOLO (You Only Look Once) family of models. The YOLOv5m6-Face model, developed by Kuzdeuov et al. [31] and employed in this study, utilizes CSPNet backbone[43] with depth and width multipliers of 0.50 and 0.75, respectively. The model underwent training for 32 epochs using the SGD optimizer. The training process incorporated an initial learning rate of 0.01, which was gradually reduced to a final rate of 0.00001. Full details on the model's architecture and training process can be found in the author's original paper [31].

The model was trained on a dataset of 16,509 labelled thermographic face images captured in various environments, including controlled indoor, semi-controlled indoor and uncontrolled outdoor [31]. The dataset used in this study is publicly available and can be accessed at <https://github.com/IS2AI/TFW>, with the permission to use, modify, and distribute under the MIT license. This dataset provides a diverse range of facial images under different environmental conditions, which is important for ensuring the robustness and generalisability of the AFFT approach. For training, a loss function was used to compare the predicted and ground truth bounding boxes and landmark locations, and the model weights were updated through backpropagation to minimize this loss. Key parameters of the model like the anchor boxes and detection thresholds were set empirically based on this dataset.

In application, the trained YOLOv5m-Face model takes a new thermographic image as input and outputs the predicted facial bounding boxes and landmark locations using the learned features and regression algorithms. The predicted landmark coordinates were then used to derive the nose tip temperature as described.

Firstly, *flirextractor* was used to extract raw thermographic data from our images, remove the spot annotations, logo, and other non-related metadata. Next, the images were converted from RGB to grayscale which were then inverted to obtain negatives of the images. As recommended by Kuzdeuov et al. [31] image brightness was enhanced through Gamma correction, which improved the visibility of facial features, especially eye and lip contours. Facial feature inference was then performed using the Google Colaboratory [[6]](#footnote-6) platform. This study utilised the Google Colab (https://colab.research.google.com) free GPU instance for conducting experiments. The required software was obtained from the original dataset/model providers GitHub repository.

### 3.6. Data Analysis Approach

The data obtained from the MTLA, TST and AFFT based approaches were all obtained from the thermographic images of the participants and were normalized to remove any unstructured and redundant data. 369 images in total were finally generated for processing. For each participant, there were 41 images processed. This represented data for a total of 9 participants, deemed usable for analysis. For each participant, the 41 images represented still frames at each sampling interval over the entire period of measurement.

To measure the strength of the linear association between the MTLA and TST, MTLA and AFFT approach, and TST and AFFT approach respectively, the correlation coefficient was calculated in Microsoft Excel with Equation (1) [44].

……………… (1)

Where the correlation coefficient is , are the variables and are their mean values of the variables. The values of inclusively are between -1 and 1.

The error was determined by subtracting the measured values of either the AFFT approach or TST – based approach from the MTLA measurement. Absolute values were then obtained to establish absolute errors in measurement of temperature, X - and Y - coordinates. The relative error was then estimated in Microsoft Excel by dividing the absolute error by the MTLA values as shown in Equation (2).

……… (2)

To determine the similarity in the coordinate data measured between AFFT - MTLA and the TST - MTLA respectively, Euclidean distance [45], [46] was used. This was to factor in the spatial relationships between pixels [45]. Here, Equation (3) was entered into Microsoft Excel and applied to the X and Y coordinate data from the AFFT and TST measurements, against the ground truth of the MTLA.

……… (3)

Where are two points in Euclidean -space, and are the initial Euclidean vectors.

## 4. Results and Discussion

### 4.1. Facial Landmark Labelling Performance

The practicality of using thermal imaging techniques in realistic working environments were evaluated. To do so, TST and AFFT, the two techniques used in determining the nose tip temperature over time, were compared to each other. The absolute error between the MTLA temperature measurement and AFFT temperature measurement, and the MTLA temperature measurement and TST temperature measurement were respectively determined. A similar determination was done for the X and Y-coordinate markers using the two approaches. To estimate the measurement accuracy, relative error was determined using equation (2) for the temperature measurement, X-coordinate position, and Y-coordinate position respectively (See Figure 9). In each Figure, the Image ID is represented by an image code/file name in the form *IR\_XXXX* for each image processed.

*9a) Relative error for temperature measurement  
  
9b) Relative error for x-coordinate position  
  
9c) Relative error for y-coordinate position*

In Figure 9a, the plot shows that the relative error in temperature between the MTLA and the TST was comparatively higher than that between the MTLA and the AFFT approach. Here the mean relative error in temperature between the MTLA and the TST was 0.05°C while that between the MTLA and the AFFT was about 0.06°C. The corresponding maximum errors in temperature were respectively 0.28°C and 0.3°C respectively. It is important to recognize that for facial temperature measurements, slight increments may have different and significant implications. Here, a standard deviation of 0.07°C was determined for the relative error involving the AFFT as opposed to a standard deviation of 0.06°C determined for the relative error involving the TST.

The error associated with the temperature measurements using the two automated approaches were because of both techniques capacity to pick out X and Y-coordinates representative of the nose tips of participants. In Figure 9b, the relative error in x-coordinate position between the MTLA nose tip measurement and AFFT nose tip measurement, and the MTLA nose tip measurement and TST nose tip measurement are presented. Between the MTLA and the TST, the average and maximum relative errors were respectively 0.02 and 0.09. For the MTLA and the AFFT approach, the average and maximum relative errors were respectively 0.009 and 0.04. Here, the results show that the AFFT had a better accuracy in selecting the x-coordinate of the participants nose tips to that of the TST approach. The standard deviation of the relative error data for the MTLA/AFFT was 0.008 and that for the MTLA/TST was 0.01.

Figure 9c presents superimposed plots of the relative error in Y-coordinate position between the MTLA nose tip measurement and AFFT nose tip measurement, and the MTLA nose tip measurement and TST nose tip measurement. Here, the average and maximum relative errors of the Y-coordinate between MTLA nose tip measurement and TST nose tip measurement were respectively 0.04 and 0.21. While for the MTLA nose tip measurement and AFFT nose tip measurement, they were respectively 0.01 and 0.08. The standard deviation of the relative error data for the Y-coordinate involving the MTLA and AFFT approach was 0.01 and that for the MTLA/TST was 0.03.

Euclidean distance was employed to compute the disparities between the coordinate data points of MTLA\_x/AFFT\_x and MTLA\_y/AFFT\_y, as well as those of MTLA\_x/TST\_x and MTLA\_y/TST\_y. The outcomes of these computations are illustrated in Figure 10. The mean distance for the AFFT method was determined to be 5.4, whilst the mean distance for the TST method was 15.3. In terms of maximum distance, the AFFT method exhibited a peak of 26.6, whereas the maximum distance for the TST method was 59.9. The standard deviation for the AFFT method was calculated to be 4.63, whilst the TST method had a standard deviation of 7.01. On the basis of these results, it can be concluded that the AFFT method demonstrates a higher degree of accuracy and precision in comparison to the TST method.

*Figure 10 Euclidean distance between the ground truth (MTLA) and the automatic (AFFT) and spot-based (TST) approaches.*

### 4.2. Coordinates Mapping

The nose tip temperature data was obtained by averaging the temperature values from the neighbouring pixels at the X and Y coordinates representing the nose tip location in the infrared images of the participants. With the MTLA involving the manual annotation using the data labelling software RectLabel1, the nose tip of each participant was manually selected as shown in Figures 7a and 7b. Each nose tip selected represented a particular X and Y coordinate that allowed the temperature to be read off. For the TST approach, facial spots including the nose tip were preselected from the settings of the FLIR T640 thermal imaging camera for each participant. The temperature values reported directly by the FLIR camera for these spots were used without averaging from neighbouring pixels. For the AFFT approach, the AFFT was trained to detect the nose tip to obtain the relevant temperature for each case study participant. As shown in Figures 11, 12 and 13, although the same nose tip temperature of each participant was to be obtained, there was variation in accuracy. This was largely due to the movement of participants in the case of the TST technique and in the case of the AFFT it was because of prediction accuracy of the trained model.

In Figure 11, X-coordinate positions for the nose tips of the infrared images were compared for the three approaches. It is important to recognize that each participant had different facial structure hence the manual adjustment of the spots in the FLIR T640 thermal imaging camera settings changed for each measurement. However, based on the MTLA providing the actual reference coordinate, the AFFT approach and the TST X-coordinate positions were compared to it to see how far they were from the MTLA. The correlation coefficient, between the MTLA and the AFFT approach was determined to be 0.93 while between the MTLA and the TST it was determined to be 0.81. These results showed a positive correlation coefficient between the MTLA and the AFFT approach and a positive correlation coefficient between the MTLA and the TST technique. As can also be seen in Figure 11, the camera’s X - coordinate varied minimally due to similar facial structures of the participants leading to relatively static X-coordinates compared to that of the MTLA and AFFT measurements. The MTLA however shows a varied pattern indicative of the facial structures of the participants. With that as the reference, it can be seen in Figure 11 that the AFFT at times overestimated or underestimated the relevant X-coordinate. The MTLA follows a similar trend as the TST technique but got lost when participants moved.

*Figure 11 x-coordinate positions for the nose tips of the infrared images*

Similarly, for the Y - coordinates, the correlation coefficients were determined between the MTLA - AFFT and the MTLA - TST respectively. Between the MTLA and the AFFT technique, the correlation coefficient obtained was 0.98 while between the MTLA and the TST technique it was 0.87. Here too, there was a positive correlation coefficient between the MTLA and the AFFT approach and a positive correlation coefficient between the MTLA and the TST technique. Once again, due to similar facial features, the variation of the Y-coordinates using the TST techniques was comparatively minimal as shown in the relative straight lines for the TST technique in Figure 12. This was because as the coordinates were set in the TST, those static points were maintained throughout the measurements. However, with the MTLA selecting near accurate coordinates, it roughly followed the trend of the TST approach albeit accurately as it factored in the movements of participants during the measurements.

*Figure 12 y-coordinate positions for the nose tips of the infrared images*

Figure 13 shows a superimposed scatter plot of the X and Y coordinates for the three techniques used. With the MTLA as the reference, there were fewer TST scatter points, and they were mainly in the region of the MTLA. For the AFFT detected coordinates, a good amount was in the zone of the MTLA, however there were some which were found to be off the zone of the MTLA. Although the AFFT was expected to have the capacity to accurately track the nose tip and obtain the temperature accordingly, it appeared from the results that the TST was comparatively closer to the MTLA. This is evident from the R2 values determined from the trendlines of all three techniques. From the trendlines, all three techniques showed a negative gradient. A video of the X and Y coordinates for the three approaches superimposed on participants’ thermographic images can be found here [Participants' Coordinates Mapping Video.mp4](file:///D:\KH026800\PhD\Papers\Publications\Mental%20Workload%20Paper\Participants'%20Coordinates%20Mapping%20Video.mp4) .

*Figure 13 Scatter plots for the x and y coordinates of the three techniques superimposed*

The fewer coordinate points for the TST indicates the fact that within the camera settings those coordinates were locked in for each participant to capture their relevant facial points hence when they moved the temperature of the spot detected for the coordinates was captured. The variation in coordinates here was because of the variation in facial structures of the participants. Whereas for the MTLA and the AFFT approach, there was an attempt to track each participants’ nose tip and as they moved the coordinates changed leading to several points in the scatter plot in Figure 13. The coefficient of determination (R2) values for the MTLA, AFFT, and TST in terms of the scatter of the x and y coordinates were 0.65 (65%), 0.53 (53%) and 0.48 (48%) respectively. These R2 values clearly show stronger correlation with the MTLA followed by the AFFT approach with the TST showing a weak correlation of less than 50%. The AFFT approach follows with a 11% difference in accuracy.

### 4.3. Variation in Measured Temperature

Using the three techniques, namely MTLA, AFFT and TST, the average of the nose tip temperatures obtained for all the IR images were 29.93°C, 29.95°C and 31.06°C respectively. The difference in nose tip temperature between the MTLA average and the AFFT approach average was approximately 0.02°C. For the TST, a nose tip temperature difference of 1.13°C was obtained between its average temperature and the average temperature of the MTLA dataset. This difference in nose tip temperature is crucial as in MWL measurements, it is expected that nose tip temperature decreases with increasing MWL [7], [8] and for human skin a degree in temperature change may be significant. From the same thermographic images, the MTLA, AFFT and TST obtained maximum nose tip temperatures of 34.57°C, 35.14°C, and 35.20°C respectively. It is important to note that the movement of the participants were crucial here and impacted the nose tip temperatures obtained via the AFFT and the TST approaches respectively. The TST being an inherent flaw in the FLIR T640 thermal imaging camera due to its inability to track the participants led to temperature on the other facial points being picked up for the nose tip temperature as they were static within the camera settings. This is evident from Figure 4 and [Participants' Coordinates Mapping Video.mp4](file:///D:\KH026800\PhD\Papers\Publications\Mental%20Workload%20Paper\Participants'%20Coordinates%20Mapping%20Video.mp4). of the tracking techniques where movements of participants resulted in different facial point temperatures being recorded to represent the nose tip temperature. For the AFFT approach, its capacity relied on the number and types of images the AI was trained on. The dataset used to train the neural network contained subjects in both indoor and outdoor settings. The original dataset was collected using the FLIR T540 with a resolution of 464 x 348, which may contribute to the performance differences observed here. Additionally, whilst image pre-processing techniques used in this study followed those documented in the original author’s work, it is possible that variations in the degree of processing and selected settings may contribute to performance differences also. Finally, it is difficult to know whether factors such as the race, age and sex of the original data collection contained a representative sample of the participants who took part in this study [47], [48].

In terms of the minimum nose tip temperature obtained, they were respectively 24.99°C, 22.22°C and 26.20°C for the MTLA, AFFT approach and TST. The key thing is that due to the possibility of erroneous temperature data because of participants’ movement, this cannot be easily attributed to the nose tip temperature only for the AFFT approach and TST approach as can be seen in the [Participants' Coordinates Mapping Video.mp4](file:///D:\KH026800\PhD\Papers\Publications\Mental%20Workload%20Paper\Participants'%20Coordinates%20Mapping%20Video.mp4). This is unlike the MTLA where the nose tip coordinates were selected to obtain the exact measured nose tip temperature by the FLIR T640 thermal imaging camera. The standard deviations of the temperature data for the MTLA, AFFT and TST techniques were 3.30°C, 3.37°C and 2.43°C. Here, there was a comparative difference between the MTLA and the TST measurements, a 0.87°C difference. Comparative between the MTLA and the AFFT, there was not much difference, a 0.07°C difference. Once again, it is important to note that the averages of the AFFT and TST techniques included temperature data that were not exactly on the nose tips of the volunteering participants as can be seen in the [Participants' Coordinates Mapping Video.mp4](file:///D:\KH026800\PhD\Papers\Publications\Mental%20Workload%20Paper\Participants'%20Coordinates%20Mapping%20Video.mp4). As can be seen in Figure 14, the maximum nose tip temperature range, between the minimum and the maximum for each approach was roughly around 9°C. This is because in the case study focused on assessing MWL, the range of temperatures that can be recorded for the face was not of a significantly large margin.

*Figure 14 Nose tip temperatures obtained from all the infrared images*

### 4.4. Measurement Uncertainty

The facial temperature measurements were taken with the FLIR T640 thermal imaging camera, while the environmental conditions were monitored with Omega K-type thermocouples and AZ 8829 sensor and data logger. Table 1 presents the absolute uncertainty values for the measured fundamental parameters attesting to the accuracy of the measurements.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1 Absolute Uncertainty Values for Fundamental Parameters** | | | |
| **Parameter** | **Measurement Device** | **Absolute Uncertainty** | **Units** |
| Facial Temperature (1.4m from participants) – Primary Camera | FLIR T640 thermal imaging camera | Accuracy +/-2% or 2°C  Thermal Sensitivity <0.04°C  Temperature range (-40°C to 2,000°C) | °C |
| Facial Temperature (1.4m from participants) – Secondary Camera | FLIR E75 thermal imaging camera | Accuracy ±2°C or ±2% of reading, for ambient temperature 15 to 35°C and object temperature above 0°C | °C |
| Ambient Stratified Temperature | Omega K Type Thermocouples | (±2.2C or 0.75%) for standard limits of error (per ANSI/ASTM E230) | °C |
| Ambient Temperature (Desktop level) | AZ 8829 sensor and data logger | ±0.6 (from -20~50°C), ±1.2 (others) | °C |
| Relative Humidity (Desktop level) | AZ 8829 sensor and data logger | (Humidity Resolution ±0.1)  Accuracy ±3% | % |

### 4.5. Valuable Learnings from the Study

There were several valuable learnings, which are presented concisely in the title of this paper as "What not to do" when conducting such studies. These lessons are:

* + **Inconsistent Camera Panning/Zooming between subjects**: One of the main challenges we faced during our study was inconsistent camera panning and zooming between subjects while capturing thermal images. The undergraduate researcher who performed the imaging sessions unintentionally varied the level of panning and zooming between subjects, without recording the degree to which these had been performed. Our processing pipeline was not designed to manage these variations, and seemingly this information was not preserved in each image's metadata, leading us to exclude these subjects' data in our analysis. In addressing this issue, we recommend implementing a standardised protocol for camera operation to ensure consistency across all subjects. This could involve providing additional or specific training/guidelines to researchers on this aspect of the camera's configuration; or/as-well-as using software that can automatically adjust for camera movements during image capture.
  + **Subject Movement Impact Accuracy**: The primary objective of using AFFT is to eliminate restrictions on participant movement during imaging sessions. However, there are still limitations that may affect the accuracy of predicting facial features. Camera occlusion and quick head-movement can interfere with current models' ability to accurately identify facial landmarks. Improved facial landmark identification can be achieved by enhancing AFFT models to better predict facial features and developing more advanced algorithms that utilize larger datasets. Additionally, incorporating generative AI models such as those presented in Batchuluun et al., [49] could further improve this development. Furthermore, providing additional training for researchers on how to optimize camera positioning and movement to minimize interference would also aid in addressing this issue.
  + **Manual Data Labelling**: The foundation of our work is based on the use of manually annotated thermal image data to establish a ground-truth dataset. However, this process was challenging due to several factors, including time consumption and ambiguity in certain images regarding facial feature locations. This issue often occurred when subjects' heads turned slightly. One advantage of our experimental setup was having a clear temperature difference between participants and their surroundings, as we had a blank unobstructed wall directly behind them. In more realistic studies, however, this may not always be possible, which could further affect data label quality. High-quality labelled data is crucial for establishing accurate ground truths and providing high-quality datasets for training future DL models. One interesting approach to addressing this challenge could be to integrate an AI into the data labelling process, whereby an AI suggests possible placement of key facial features which are then reviewed by the (human) researcher. Similar approaches have been applied in the labelling of 3D medical images in Diaz-Pinto et al., [50].
  + **Data Processing in Real World applications**: While our study did not specifically explore the use of this approach in real-world, real-time applications, it is a promising area for future work. However, we acknowledge that the resource requirements for data processing were substantial and exceeded what would be feasible in most real-world settings. To ensure that this method remains accessible to researchers, efforts must be made to optimise models without relying on very high-end computing hardware. This may involve developing more efficient algorithms or implementing cloud-based solutions that can manage large datasets with lower computational demands.

### 4.6. Study Limitations and Future Work

The way the camera was operated during the data-collection phase of the study presented some challenges. For each participant, the student researcher would manually zoom and pan the thermographic camera to ensure that the participants face featured prominently in the viewfinder. Our thermal imaging camera, however, only supports digital zoom, meaning that this functionality manifested, essentially as a cropping of the higher quality image. Using the technical approach described here, we were able to recover the original (non-zoomed) image to identify the zoom-factor, from which we hoped to be able to re-locate the original spots, by accounting for the original zoom-in. In addition to zooming in upon subjects’ faces, the student researcher also manually panned the zooming window to centre the participants nose using the crosshair overlay. The degree to which the student researcher panned is not, to the best of our knowledge, recorded in the metadata reported by *exiftool*, meaning we were unable to account for the panning in our data processing.

We were however able to benchmark the distance performance of our AFFT (AI-tracking) approach since this works off the original (non-zoomed) image. For our 640x480 (307,200) pixel images, our AFFT approach was, on average, within 27 pixels (STD: 42.4 pixels) of the annotation recorded in the MTLA dataset.

It is important to note that the analysis presented here targets minimizing measurement errors using the FLIR T640 thermal imaging camera based on the human error aspects and not the inherent measurement limitations of the camera. This means that the temperature measurement results obtained were based on the sensitivity of the thermal camera and its inherent accuracy or uncertainty as presented in Table 1.

The MTLA approach though comparatively more accurate is a time consuming and laborious process, making it unsuitable for use under certain task types, working environments or where a large amount of data needs to be processed and analysed. This means its effectiveness is in its use on a reference basis and not as the main approach to annotate and obtain large amounts of data. This is because in a small dataset it may be practical, however where there is big data or situations where there are significant movements by subjects or targets, automation with AFFT approach is likely to be more useful.

The study's reliance on a single/existing dataset [30] may also pose limitations related to data diversity and generalisability. It is important to consider the integrity of the data used to train the ML system, and new attack vectors such as systematic poisoning attacks should be taken into account [51]. Future work should aim to incorporate diverse datasets to improve the robustness and generalisability of the model.

We did not explore additional contexts for deploying this approach, such as the healthcare sector, which presents clear application opportunities[52]. Future work should investigate the feasibility of using this approach in healthcare settings, but this will necessitate additional considerations, especially related to data security and privacy. With the sensitivity of data in healthcare, care must be taken to consider data encryption and security [53], [54]. The evolving landscape of post-quantum cryptographic hardware and accelerators may render existing encryption methods, such as AES, obsolete[55]. Therefore, it is essential to consider how these technologies will be deployed, particularly in low-power platforms that could benefit less developed areas [56], [57].

For future work, integrating additional noise reduction techniques and image normalisation into the preprocessing stage can help ensure consistent thermal measurements [58]. Incorporating multi-view geometry[59] may also enhance landmark detection accuracy by considering different angles and perspectives. Exploring more advanced deep learning models such as Faster R-CNN[60] or HRNet [61] or fine-tuning existing models using transfer learning from pre-trained models on large facial recognition datasets [59] are other approaches to improve performance.

## 5. Conclusion

In this study, we conducted a comprehensive comparison of the AFFT approach against two other methods (MTLA and TST) utilising various metrics such as temperature measurement accuracy, coordinate mapping accuracy, and measurement uncertainty, allowing us to establish the relative strengths and weaknesses of each approach. The AFFT was based on a model trained on a large dataset of thermographic images using a modern deep learning neural network (CNN) based approach (YOLO). Additionally, the study identified and addressed several sources of error in thermal imaging measurements, improving the accuracy of the measurements and the reliability of the results. Our findings provide a method for realising applications of this approach in future research and practical implementations. Our findings show that:

* + The mean relative error between the MTLA and TST methods was 0.05°C, while that between MTLA and AFFT was about 0.063°C. Maximum errors were 0.28°C for MTLA-AFFT and 0.3°C for TST-MTLA, indicating that AFFT provided temperatures closer to MTLA's measurement.
  + The mean relative errors for X and Y-coordinate positions were 0.02 and 0.04 respectively between MTLA and TST, with maximum errors of 0.09 and 0.21. Between MTLA and AFFT, the mean relative errors were 0.009 and 0.01 respectively, with corresponding maximum errors of 0.04 and 0.08.
  + The R2 values for nose tip position measurements using the MTLA, AFFT, and TST methods were 0.65, 0.53, and 0.42 respectively. These R2 values indicate that the MTLA had a stronger correlation than the other two approaches, with the Spot-based method showing the weakest correlation at less than 50%.
  + The AFFT method outperforms the TST method in terms of accuracy and precision, with a mean Euclidean distance of 5.4 versus 15.3, a maximum distance of 26.6 versus 59.9, and a standard deviation of 4.63 versus 7.02 for coordinate data point disparities.

Our future research will focus on enhancing AFFT to work in more diverse and dynamic environments, especially in realistic living environments such as offices and homes. While this study demonstrates the potential of this approach, we acknowledge that there is room for improving its accuracy and applicability. One way towards achieving this goal is through the development of more sophisticated DL models, trained on a larger dataset than used in this work - an endeavour we are currently undertaking. Furthermore, it is important to note that the results presented in this study were derived from thermal images sourced from a single FLIR T640 imager, which will not be representative of all thermal imagers. To fully realise the potential of our work, future studies must incorporate images from a variety of different thermal imagers, each with their own unique performance characteristics. In addition, advanced DL models will aim to mitigate movement and other artifact noise impacting image quality and model performance. There is existing work, such as work by Batchuluun et al., [49], that provides approaches towards resolving these issues, which we plan to include in future iterations of this work. The applications areas for this work are broad and not limited to any particular discipline or industry - anything involving human-interaction can benefit from the lighting-independent, privacy-preserving, and physiologically derived insights provided by this measure.

## 6. Ethics Statement

This study, exploring the relationship between participant's nose tip temperature change and Mental Workload (MWL) in a realistic (uncontrolled) working environment was reviewed and approved by the University of Nottingham Ningbo China’s (UNNC) Research Ethics Sub-Committee. Each participant was provided with a written standard informed consent form, approved by the UNNC Research Ethics Sub-Committee, to sign before partaking in the measurements.

## 7. Acknowledgements

The authors would like to thank Ms Kate Yuan of the Faculty of Science and Engineering (FoSE) of the University of Nottingham Ningbo China (UNNC) for her support in helping set up the equipment for the monitoring of the environment of the measurement location. We are also deeply indebted to all the volunteering participants who were all students of UNNC at the time for their participation in the project.

## 8. References

[1] S. Midha, H. A. Maior, M. L. Wilson, and S. Sharples, “Measuring Mental Workload Variations in Office Work Tasks using fNIRS,” *Int J Hum Comput Stud*, vol. 147, p. 102580, Mar. 2021, doi: 10.1016/j.ijhcs.2020.102580.

[2] M. F. Pike, H. A. Maior, M. Porcheron, S. C. Sharples, and M. L. Wilson, “Measuring the effect of think aloud protocols on workload using fNIRS,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, Apr. 2014, pp. 3807–3816. doi: 10.1145/2556288.2556974.

[3] H. A. Maior, M. Pike, S. Sharples, and M. L. Wilson, “Examining the Reliability of Using fNIRS in Realistic HCI Settings for Spatial and Verbal Tasks,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, Apr. 2015, pp. 3039–3042. doi: 10.1145/2702123.2702315.

[4] C. Berka *et al.*, “EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks.,” *Aviat Space Environ Med*, vol. 78, no. 5 Suppl, pp. B231-44, May 2007.

[5] P. Nickel and F. Nachreiner, “Sensitivity and Diagnosticity of the 0.1-Hz Component of Heart Rate Variability as an Indicator of Mental Workload,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 4, pp. 575–590, Dec. 2003, doi: 10.1518/hfes.45.4.575.27094.

[6] C. D. Wickens, “Multiple Resources and Mental Workload,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 3, pp. 449–455, Jun. 2008, doi: 10.1518/001872008X288394.

[7] A. C. Marinescu, S. Sharples, A. C. Ritchie, T. Sánchez López, M. McDowell, and H. P. Morvan, “Physiological Parameter Response to Variation of Mental Workload,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 60, no. 1, pp. 31–56, Feb. 2018, doi: 10.1177/0018720817733101.

[8] K. Murai, T. Okazaki, L. C. Stone, and Y. Hayashi, “A characteristic of a navigator&#x2019;s mental workload based on nasal Temperature,” in *2007 IEEE International Conference on Systems, Man and Cybernetics*, IEEE, Oct. 2007, pp. 3639–3643. doi: 10.1109/ICSMC.2007.4413737.

[9] C. K. L. Or and V. G. Duffy, “Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement,” *Occupational Ergonomics*, vol. 7, no. 2, pp. 83–94, Jul. 2007, doi: 10.3233/OER-2007-7202.

[10] K. Schweikert, A. Sielaff, and P. Stephan, “Pixel-wise in situ calibration method for high accuracy infrared thermography of moving targets,” *Infrared Phys Technol*, vol. 118, p. 103862, Nov. 2021, doi: 10.1016/j.infrared.2021.103862.

[11] G. J. Tattersall, “Infrared thermography: A non-invasive window into thermal physiology,” *Comp Biochem Physiol A Mol Integr Physiol*, vol. 202, pp. 78–98, Dec. 2016, doi: 10.1016/j.cbpa.2016.02.022.

[12] N. Playà‐Montmany and G. J. Tattersall, “Spot size, distance and emissivity errors in field applications of infrared thermography,” *Methods Ecol Evol*, vol. 12, no. 5, pp. 828–840, May 2021, doi: 10.1111/2041-210X.13563.

[13] X. Tian, L. Fang, and W. Liu, “The influencing factors and an error correction method of the use of infrared thermography in human facial skin temperature,” *Build Environ*, vol. 244, p. 110736, Oct. 2023, doi: 10.1016/j.buildenv.2023.110736.

[14] D. J. Bell, A. I. Macrae, M. A. Mitchell, C. S. Mason, A. Jennings, and M. J. Haskell, “Comparison of thermal imaging and rectal temperature in the diagnosis of pyrexia in pre-weaned calves using on farm conditions,” *Res Vet Sci*, vol. 131, pp. 259–265, Aug. 2020, doi: 10.1016/j.rvsc.2020.05.004.

[15] L. Jiao, D. Dong, X. Zhao, and P. Han, “Compensation method for the influence of angle of view on animal temperature measurement using thermal imaging camera combined with depth image,” *J Therm Biol*, vol. 62, pp. 15–19, Dec. 2016, doi: 10.1016/j.jtherbio.2016.07.021.

[16] D. Li, C. C. Menassa, and V. R. Kamat, “Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography,” *Energy Build*, vol. 176, pp. 246–261, Oct. 2018, doi: 10.1016/j.enbuild.2018.07.025.

[17] M. Fajkus, J. Nedoma, R. Martinek, V. Vasinek, H. Nazeran, and P. Siska, “A Non-Invasive Multichannel Hybrid Fiber-Optic Sensor System for Vital Sign Monitoring,” *Sensors*, vol. 17, no. 12, p. 111, Jan. 2017, doi: 10.3390/s17010111.

[18] H. Metzmacher, D. Wölki, C. Schmidt, J. Frisch, and C. van Treeck, “Real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment,” *Energy Build*, vol. 158, pp. 1063–1078, Jan. 2018, doi: 10.1016/j.enbuild.2017.09.032.

[19] A. S. and Morris and R. Langari, *Measurement and Instrumentation : Theory and Application*, 2nd ed. Elsevier Science & Technology, 2015. Accessed: Apr. 03, 2022. [Online]. Available: https://ebookcentral.proquest.com/lib/lsbuuk/detail.action?docID=5754522

[20] J. Ranjan and J. Scott, “ThermalSense,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, NY, USA: ACM, Sep. 2016, pp. 1212–1222. doi: 10.1145/2971648.2971659.

[21] M. W. Akram, G. Li, Y. Jin, X. Chen, C. Zhu, and A. Ahmad, “Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning,” *Solar Energy*, vol. 198, pp. 175–186, Mar. 2020, doi: 10.1016/j.solener.2020.01.055.

[22] T. D’Orazio, C. Guaragnella, M. Leo, and P. Spagnolo, “Defect detection in aircraft composites by using a neural approach in the analysis of thermographic images,” *NDT & E International*, vol. 38, no. 8, pp. 665–673, Dec. 2005, doi: 10.1016/j.ndteint.2005.04.005.

[23] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun, and Y. Sun, “Infrared Thermal Imaging-Based Crack Detection Using Deep Learning,” *IEEE Access*, vol. 7, pp. 182060–182077, 2019, doi: 10.1109/ACCESS.2019.2958264.

[24] S. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, “Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model,” *Sensors*, vol. 18, no. 9, p. 2799, Aug. 2018, doi: 10.3390/s18092799.

[25] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, “DeepBreath: Deep Learning of Breathing Patterns for Automatic Stress Recognition using Low-Cost Thermal Imaging in Unconstrained Settings,” Aug. 2017, doi: 10.1109/ACII.2017.8273639.

[26] C. R. Siah, S. T. Lau, S. S. Tng, and C. H. M. Chua, “Using infrared imaging and deep learning in fit‐checking of respiratory protective devices among healthcare professionals,” *Journal of Nursing Scholarship*, vol. 54, no. 3, pp. 345–354, May 2022, doi: 10.1111/jnu.12736.

[27] Y. Wang *et al.*, “Unobtrusive and Automatic Classification of Multiple People’s Abnormal Respiratory Patterns in Real Time Using Deep Neural Network and Depth Camera,” *IEEE Internet Things J*, vol. 7, no. 9, pp. 8559–8571, Sep. 2020, doi: 10.1109/JIOT.2020.2991456.

[28] A. Akula, A. K. Shah, and R. Ghosh, “Deep learning approach for human action recognition in infrared images,” *Cogn Syst Res*, vol. 50, pp. 146–154, Aug. 2018, doi: 10.1016/j.cogsys.2018.04.002.

[29] A. Bhattacharyya, S. Chatterjee, S. Sen, A. Sinitca, D. Kaplun, and R. Sarkar, “A deep learning model for classifying human facial expressions from infrared thermal images,” *Sci Rep*, vol. 11, no. 1, p. 20696, Oct. 2021, doi: 10.1038/s41598-021-99998-z.

[30] S. D. Lin, L. Chen, and W. Chen, “Thermal face recognition under different conditions,” *BMC Bioinformatics*, vol. 22, no. S5, p. 313, Nov. 2021, doi: 10.1186/s12859-021-04228-y.

[31] A. Kuzdeuov, D. Aubakirova, D. Koishigarina, and H. A. Varol, “TFW: Annotated Thermal Faces in the Wild Dataset,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2084–2094, 2022, doi: 10.1109/TIFS.2022.3177949.

[32] B. Cinaz, B. Arnrich, R. la Marca, and G. Tröster, “Monitoring of mental workload levels during an everyday life office-work scenario,” *Pers Ubiquitous Comput*, vol. 17, no. 2, pp. 229–239, Feb. 2013, doi: 10.1007/s00779-011-0466-1.

[33] X. Sun, H. Wu, and Y. Wu, “Investigation of the relationships among temperature, illuminance and sound level, typical physiological parameters and human perceptions,” *Build Environ*, vol. 183, p. 107193, Oct. 2020, doi: 10.1016/j.buildenv.2020.107193.

[34] K. W. Tham and H. C. Willem, “Room air temperature affects occupants’ physiology, perceptions and mental alertness,” *Build Environ*, vol. 45, no. 1, pp. 40–44, Jan. 2010, doi: 10.1016/j.buildenv.2009.04.002.

[35] A. Ghahramani, G. Castro, B. Becerik-Gerber, and X. Yu, “Infrared thermography of human face for monitoring thermoregulation performance and estimating personal thermal comfort,” *Build Environ*, vol. 109, pp. 1–11, Nov. 2016, doi: 10.1016/j.buildenv.2016.09.005.

[36] Pixabay GmbH, “pixaby,” Man Head Face Avatar Low Poly Geometric - https://pixabay.com/vectors/man-head-face-avatar-low-poly-6060948/.

[37] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “Facial Landmark Detection by Deep Multi-task Learning,” 2014, pp. 94–108. doi: 10.1007/978-3-319-10599-4\_7.

[38] Y. Wu and Q. Ji, “Facial Landmark Detection: A Literature Survey,” *Int J Comput Vis*, vol. 127, no. 2, pp. 115–142, Feb. 2019, doi: 10.1007/s11263-018-1097-z.

[39] A. Klink, “aloisklink/flirextractor v1.0.0: An efficient GPLv3 Python package for extracting temperature data from FLIR IRT images,” Nov. 15, 2019, *University of Southampton, Southampton*.

[40] W. Mckinney, “pandas: a Foundational Python Library for Data Analysis and Statistics.” [Online]. Available: https://www.researchgate.net/publication/265194455

[41] D. Qi, W. Tan, Q. Yao, and J. Liu, “YOLO5Face: Why Reinventing a Face Detector,” May 2021.

[42] G. Jocher *et al.*, “ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models,  AWS, Supervise.ly and YouTube integrations,” Apr. 2021, *Zenodo*. doi: 10.5281/zenodo.4679653.

[43] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, “CSPNet: A New Backbone that can Enhance Learning Capability of CNN,” Nov. 2019.

[44] Microsoft, “CORREL function.” Accessed: Sep. 19, 2023. [Online]. Available: https://support.microsoft.com/en-gb/office/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92

[45] Liwei Wang, Yan Zhang, and Jufu Feng, “On the Euclidean distance of images,” *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 8, pp. 1334–1339, Aug. 2005, doi: 10.1109/TPAMI.2005.165.

[46] T. P. Rinjeni, A. Indriawan, and N. A. Rakhmawati, “Matching Scientific Article Titles using Cosine Similarity and Jaccard Similarity Algorithm,” *Procedia Comput Sci*, vol. 234, pp. 553–560, 2024, doi: 10.1016/j.procs.2024.03.039.

[47] D. v. Carvalho, E. M. Pereira, and J. S. Cardoso, “Machine Learning Interpretability: A Survey on Methods and Metrics,” *Electronics (Basel)*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/electronics8080832.

[48] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A Survey on Bias and Fairness in Machine Learning,” *ACM Comput Surv*, vol. 54, no. 6, pp. 1–35, Jul. 2021, doi: 10.1145/3457607.

[49] G. Batchuluun, J. K. Kang, D. T. Nguyen, T. D. Pham, M. Arsalan, and K. R. Park, “Deep Learning-Based Thermal Image Reconstruction and Object Detection,” *IEEE Access*, vol. 9, pp. 5951–5971, 2021, doi: 10.1109/ACCESS.2020.3048437.

[50] A. Diaz-Pinto *et al.*, “MONAI Label: A framework for AI-assisted Interactive Labeling of 3D Medical Images,” Mar. 2022.

[51] M. Mozaffari-Kermani, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, “Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare,” *IEEE J Biomed Health Inform*, vol. 19, no. 6, pp. 1893–1905, Nov. 2015, doi: 10.1109/JBHI.2014.2344095.

[52] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, “Deep Learning for Infrared Thermal Image Based Machine Health Monitoring,” *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 151–159, Feb. 2018, doi: 10.1109/TMECH.2017.2722479.

[53] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, “Big healthcare data: preserving security and privacy,” *J Big Data*, vol. 5, no. 1, p. 1, Dec. 2018, doi: 10.1186/s40537-017-0110-7.

[54] P. Kaur, M. Sharma, and M. Mittal, “Big Data and Machine Learning Based Secure Healthcare Framework,” *Procedia Comput Sci*, vol. 132, pp. 1049–1059, 2018, doi: 10.1016/j.procs.2018.05.020.

[55] D. J. Bernstein and T. Lange, “Post-quantum cryptography,” *Nature*, vol. 549, no. 7671, pp. 188–194, Sep. 2017, doi: 10.1038/nature23461.

[56] K. Shahbazi and S.-B. Ko, “Area and power efficient post-quantum cryptosystem for IoT resource-constrained devices,” *Microprocess Microsyst*, vol. 84, p. 104280, Jul. 2021, doi: 10.1016/j.micpro.2021.104280.

[57] G. Xu *et al.*, “PPSEB: A Postquantum Public-Key Searchable Encryption Scheme on Blockchain for E-Healthcare Scenarios,” *Security and Communication Networks*, vol. 2022, pp. 1–13, Mar. 2022, doi: 10.1155/2022/3368819.

[58] K. Liu, H. Chen, W. Bao, and J. Wang, “Thermal imaging spatial noise removal via deep image prior and step-variable total variation regularization,” *Infrared Phys Technol*, vol. 134, p. 104888, Nov. 2023, doi: 10.1016/j.infrared.2023.104888.

[59] D. Tiwari, M. Dixit, and K. Gupta, “Deep Multi-View Breast Cancer Detection: A Multi-View Concatenated Infrared Thermal Images Based Breast Cancer Detection System Using Deep Transfer Learning,” *Traitement du Signal*, vol. 38, no. 6, pp. 1699–1711, Dec. 2021, doi: 10.18280/ts.380613.

[60] A. J. Das, S. Choudhury, N. Saikia, and S. C. Rajbongshi, “Pedestrian detection in thermal and color images using a new combination of saliency network and Faster R-CNN,” *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 1239–1252, May 2023, doi: 10.12785/ijcds/1301101.

[61] K. Khabarlak and L. Koriashkina, “Fast Facial Landmark Detection and Applications: A Survey,” Jan. 2021, doi: 10.24215/16666038.22.e02.

### Appendix 1: Participant information sheet and consent form

文本, 信件

描述已自动生成

文本, 应用程序, 信件

描述已自动生成

1. School of the Built Environment and Architecture, London South Bank University, 103 Borough Road, London, SE1 0AA, United Kingdom. Email: [yeboahs5@lsbu.ac.uk](mailto:yeboahs5@lsbu.ac.uk) \* ORCID: <https://orcid.org/0000-0003-4308-9829>

   \*Corresponding Author [↑](#footnote-ref-1)
2. **Abbreviations**

   • AFFT - Automated Facial Feature Tracking

   • AI - Artificial Intelligence

   • CNN - Convolutional Neural Network

   • DL – Deep Learning

   • EEG - Electroencephalogram

   • fNIRS - functional Near Infrared Spectroscopy

   • FT - Facial Thermography

   • HRV - Heart-Rate Variability

   • ML – Machine Learning

   • MTLA - Manual Thermal Landmark Annotation

   • MWL - Mental Workload

   • TST - Thermal Spot Tracking [↑](#footnote-ref-2)
3. RectalLabel Annotation Software: <https://rectlabel.com> [↑](#footnote-ref-3)
4. flirextractor Python Library: <https://github.com/aloisklink/flirextractor> [↑](#footnote-ref-4)
5. exiftool: <https://exiftool.org> [↑](#footnote-ref-5)
6. Google Colaboratory: <https://colab.research.google.com> [↑](#footnote-ref-6)