Non-Contact Based Modeling of Enervation

Kais Riani[1*], Salem Sharak[1*], Mohamed Abouelenien[1], Mihai Burzo[1], Rada Mihalcea[1],
John Elson[2], Clay Maranville[2], Kwaku Prakah-Asante[2], and Waqas Manzoor[2]
[1] University of Michigan, USA, [2] Ford Motor Company, USA

2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG) | 979-8-3503-4544-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/FG57933.2023.10042529

***Abstract*— Significant research is currently carried out with a focus on autonomous vehicles; research is starting to focus on areas such as the modeling of occupant states and behavioral elements. This paper contributes to this line of research by developing a pipeline that extracts physiological signals from thermal imagery and modeling occupant enervation using a fully non-contact based approach. These signals are obtained via a multimodal dataset of 36 subjects across multiple channels, including the thermal and physiological modalities. Moreover, we provide a comparative analysis of non-contact and contact based channels to model the enervation state of individuals. Our analysis indicates that non-contact physiological signals extracted from thermal imagery can reach and exceed the performance of contact-based physiological signals. In addition, modeling of enervation is possible using said non-contact physiological signals and thermal features, with an accuracy of up to 70% in identifying energized and enervated occupant states. Our findings provide a novel approach for future research and opens the possibility for integration of unrestrictive sensors in future automobiles.**

I. INTRODUCTION

Autonomous-vehicles and the technologies associated with them represent a significant growth vector in the automobile and technology sectors, with global investment in the area exceeding $100 billion [22]. Among the associated technologies, one area that has not yet been well explored is the study of the state and behaviors of the occupants of autonomous vehicles. The importance of such study can be highlighted by the opportunity to develop vehicles that are able to seamlessly accommodate the needs of its occupants, including their comfort and well-being.

Of the occupant states that is of value to monitor is the circadian rhythm, as it can provide insight into the enervation of the occupant, as well as any resulting ill effects [35]. In [37], it was stated that the circadian rhythm refers to biological variations or rhythms with a cycle of approximately 24 hours that will persist even when the organism is placed in an environment devoid of time cues, while the Center for Disease Control and Prevention (CDC) states that circadian rhythms are internally driven cycles that rise and fall during a 24-hour period, helping one fall asleep at night and wake up in the morning. In building a more complete interpretation of occupants' states, the autonomous vehicle could stimulate occupant wakefulness through prompts, cues, or conversation. In addition, by coming to a better understanding of the occupants' associated physiological and visual cues, it could maximize cabin comfort with adjustments in lighting, audio, and driving style to cater for a sleeping child or provide rest over long trips. Works including [13] and [28] have studied the identification of circadian states, but have largely done so within the domains of psychology, medicine, and the study of sleep disorders.

Detection of circadian rhythm is done primarily via contact-based methods. For example, [13] utilized thermal and skin conductance sensors, as well as an ingested radiotelemetry pill while [8] utilized wrist-worn actigraphy, polysomnography, and direct collection of saliva samples. In addition, other works which aimed to directly detect human behaviors such as alertness often utilize contact-based methods [9]. These contact-based approaches might not be feasible to apply on drivers in vehicles and would not be readily accepted by drivers in the real-world due to the impediment, discomfort, and hassle involved with their utilization.

Motivated by the aforementioned challenges, we present this paper with four main contributions. These include:

- Using a dataset consisting of 36 subjects with thermal, audiovisual, physiological, and survey data. Recordings consisted of a baseline recording, in addition to two additional recordings, in line with previous studies such as [12] and similar to [2].
- An unsupervised approach to extract physiological signals from thermal imagery via our proposed pipeline.
- A non-contact based classification system that utilizes thermal features as well as the aforementioned extracted physiological signals to model an individual's enervation.
- A comparison of non-contact and contact-based channels for modeling the enervation state.

II. RELATED WORK

Numerous researchers investigated the use of thermal imaging in detecting human behaviors. Several studies used the thermal modality to detect alertness [20], [34]. Others have explored its potential in detecting distraction [25], [26]. On the other hand, contact-based physiological measurements have traditionally been employed to analyze human behavior and emotion [9], [10]. However, the fact that these sensors need to be connected to individuals made their usage limited in practical situations.

Several studies have attempted to extract physiological signals from thermal images as an alternative to physiological sensors [19]. The work by Sun et al. [32], presented a method for extracting the pulse by using a Fast Fourier transform

* These authors contributed equally to this work.

979-8-3503-4544-5/23/$31.00 ©2023 IEEE

(FFT) at several points along a blood vessel to isolate the thermal propagation component. In the study by Gault et al. [15] they introduced an improvement to previous work by applying wavelet based filtering instead of the FFT analysis. The work by Fei et al. [11] used wavelet analysis in order to extract the respiration signal. The authors in [33] analyzed the impact of exercising on the skin temperature using thermal images. More recently Bennett et al. [5] present a comparison between temperature based-methods and motion-based methods in extracting respiration rate from thermal video. However, research in this area has been limited and has generally focused on one signal at a time.

Despite using these modalities to model different human behaviors, a very limited number of researchers explored circadian rhythm [27]. Existing work in this area mainly focused on contact-based physiological data analysis [6], [23].

In the 2018 study by Koichi Fujiwara et al. [14], Heart rate variability (HRV), which is the RR interval (RRI) fluctuation in an ECG, measurements were gathered from 34 subjects, and sleep onsets were evaluated using electroencephalography (EEG) data by a sleep specialist. The results showed the efficiency of a heart rate variability-based anomaly detection system, which might be expanded to detect drowsiness as well as predict epileptic seizures. The work published in 2019 by Stone et al. [31], used an ambulatory wrist-worn blue light irradiance and skin temperature in addition to a generalized neural network approach in order to allow for the prediction of the circadian phase in a real-world environment. More recently, Masuda et al. [24] employed a smart wear garment to estimate the value of time and heart rate (HR) to reach the lowest point in the circadian rhythm by measuring electrocardiogram (ECG) during sleep. The approach has shown promising potential in determining the effects of jet lag on an individual's circadian rhythm. Kaduk et al. [21] established a theoretical foundation for integrating circadian rhythmicity studies into driver's state monitoring. They demonstrated the significance of the circadian state in system design. In a recent work by Cheng et al. [8], the authors used wrist actigraphs to predict dim light melatonin onset (DLMO) in fixed night shift workers.

Classification of circadian rhythm from two data points, in this case early and late periods of the day, was previously performed by [12]. In their study, they utilized the daily rate of change in the timing of the peak metabolite aMT6s between two urine collections. Others still evaluated circadian state based on clock gene expression gathered from hair follicles at three different points in the day [2].

III. Dataset

We collected data from 36 people of various ethnicities for our experiments, with each person partaking in five recordings in a smoke, alcohol, and drug-free state to model their circadian rhythm. The dataset includes 24 males and 12 females ranging in age from 18 to 32 years old and from various demographic backgrounds. Aside from thermal, audiovisual, and physiological data, 10 surveys were collected at various stages throughout the study that cover sleep patterns, demographics, personality and behavior, including the Karolinska Sleepiness Scale survey (KSS), which is a one-question survey that scores the level of sleepiness at the time of recording [3]. In addition to a baseline recording obtained on an earlier day, there were two primary recording sessions, one in the morning and one in the evening with two recordings in each session, as explained below.

A. Instruments

An enclosed recording station was employed to simulate the surroundings of a vehicle. The following instruments were used to capture our multimodal dataset during each recording which consisted of visual, acoustic, thermal, physiological, and linguistic modalities. However, for the purpose of this paper only the thermal and physiological modalities were analyzed.

- Logitech HD web camera recording the subject' upper body from an elevated angle.
- RGB Raspberry-Pi camera recording with a close-up face view.
- FLIR One consumer-grade thermal camera, recording the face of the subject.
- FLIR SC6700 thermal camera, recording the subject's face at 100 fps, with a resolution of 640x512 pixels and 7.2M electrons.
- Four Thought Technology Ltd. physiological sensors: Blood Volume Pulse (BVP) Sensor, Skin Temperature Sensor, Skin Conductance Flex/Pro Sensor, and Respiration Rate Sensor.
- A microphone is used to record the speech.

B. Scenarios

The individuals were instructed to arrange the first recording (baseline) at least three days before the other recordings in order to record their baseline data. Each participant completed sleeps surveys, such as the Karolinska Sleep Questionnaire (KSQ), Munich Chronotype Questionnaire (MCQ), and the Morningness-Eveningness Questionnaire (MEQ) Questionnaires, as well as the Drug and Drinking Survey prior to beginning the two-minute baseline recording using our system of cameras and sensors. During the recording, participants were asked to sit quietly for two minutes and breathe normally. Following this, they were requested to complete the Big Five Inventory (BFI) personality survey and the Demographic Survey.

Following the baseline recording by at least three days, we held the morning and evening sessions. The participants were instructed not to consume any caffeinated products on the day of the recordings or the night before. One session occurred in the morning, between 8 and 11 a.m., with a few cases occurring around noon. In all cases, we requested that the participants have their first session of the recordings within one hour of waking up. The second session took place later in the day, between 4 and 8 p.m., usually before going home, with one case taking place between 10 and 11 p.m.,

depending on when they woke up that day. The subjects were not permitted to sleep in between the two sessions.

Our assumption is that the individuals would be energized after waking up, whereas the control scenario in [4] suggests that they would become enervated in the evening after a long busy day. Each session lasted around 20 minutes and included two recordings: Silent and Active. The Silent recording consisted of two minutes where the subject was asked to sit still and breathe naturally while staring at an image reflecting the time of the recording. On the other hand, the Active recording was five minutes long in which the subjects were asked to speak freely about a topic of their choice. During this time, the team in charge of the recordings left the lab to allow the participants to talk freely.

At the end of each of the morning and evening sessions, the participants were invited to complete new surveys. Participants filled the KSS and Profile of Mood States 40 (POMS40) surveys at the end of the morning session, while the Pittsburgh Sleep Quality Index (PSQI), POMS40, and KSS surveys, as well as the open response, were completed during the evening session. The participants showed no signs of survey fatigue.

IV. Methodology

First, we will describe the contact based features that were extracted from the physiological sensors in the next subsection. Then we will describe the process of extracting the thermal features as well as the non-contact (NC) based physiological features from the thermal images in the following subsections. Finally, we will describe how these three different sets of features were utilized in order to detect subjects' enervation.

A. Contact Based Physiological Features

For our experiments, we processed the Blood Volume Pulse (BVP), Skin Temperature, and Respiration Rate signals, which were recorded using the Thought Technology Ltd hardware in order to extract statistical features. The respiration rate sensor was attached to the subject's torso, while the skin temperature sensor was fastened around the pinky finger and the heart rate sensor was attached to the index. BVP was sampled at a rate of 2048 Hz, whereas the other two signals were sampled at a true rate of 256 Hz, then upscaled to 2048 Hz to preserve consistency throughout the sensor suite.

The BVP features consist of time domain statistical features such as maximum, minimum, and mean, in addition to features representing the relation between IBI, NN, and pNN, which describe patterns in the interval between two normal heartbeats. Additionally, further distinct sets of statistical features that explain the spectral power statistics for very-low, low, and high frequency bands were computed. This resulted in a total of 49 derived features from the BVP signal.

From each of the two sensors, a series of six time-domain statistical features were computed. An additional four features expressing BVP and Respiration Rate statistical patterns were computed. Following the completion of feature extraction, the average value for each feature for a given subject's recording had been calculated to represent it as a single feature vector. These vectors represent the contact based physiological features that will be used in our experiments.

B. Thermal Region of Interest Identification & Tracking

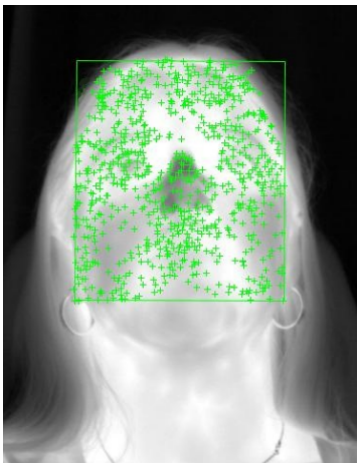


Fig. 1: Points of interest detected in the face region

We processed the thermal videos following three main steps. The first step consists of segmenting the thermal image into five different regions: the whole face, the forehead, the eyes, the cheeks, and the nose. Following that, we automatically tracked these five regions throughout the recording [30]. Finally, we created a thermal map for each region of interest (ROI) by extracting statistical features [1].

The five ROIs were manually determined in the first frame of each video recording. Using a variation of the Shi-Tomasi corner detection algorithm [29], we then automatically detected points of interest in the detected ROIs. These points suggest the existence of a blood vein affecting the temperature of the surrounding region. Fig. 1 depicts the points of interest found in the face region while utilizing a lower threshold, which allows for more points to be detected.

To stabilize the ROI bounding box, a fast version of the Kanade–Lucas–Tomasi (KLT) tracking method [30] was used on the detected points during the duration of the videos. Tracking of the points of interest was accomplished by measuring the displacement between two subsequent frames. Following that, we used geometric transformation [18] to map the important points between the frames by predicting their transformation based on similarity. To accommodate for probable occlusion, we established a threshold of 95% of properly mapped points between two successive frames as a precaution. When an occlusion was present, the frame was skipped, and tracking resumed at the point when the occlusion terminated. Finally, for each ROI, we generated a thermal map that represented the thermal distribution. This was accomplished through the procedures of ROI segmentation, segment binarization, image masking, and finally thermal map cropping [27]. The final feature vector that represented

each recording consisted of 125 features, including a 20-bin histogram, maximum, minimum, range, mean, and the mean of the top highest 10% of temperatures in the region.

These vectors represent the thermal features that were used in our experiments. In addition, the ROI tracking process is beneficial to the NC physiological features extraction, as described in the following subsections.

C. Non-Contact Physiological Features

In addition to thermal feature vectors, we also extracted the three aforementioned physiological signals from the thermal videos to provide an NC alternative. In order to extract the respiration rate signal from thermal images, we chose the maxillary (nose) region as the region of interest, as it clearly shows the subject's breathing rate with cool air entering and warm air exiting the nostrils. For the heart rate, we chose the ocular (eyes) region and the forehead region as a whole, as well as the blood vessels located in the forehead and the inner corners of the eyes (inner canthus). In the next subsections we describe the latter two in further detail. In all cases, the raw thermal signals were created by averaging the pixel values inside the zones of interest for each video frame. After producing the feature vectors of the selected ROI, we processed the data to extract the relevant signal. Our signal extraction pipeline consists of six steps, namely: Differencing, Normalization, Downsampling, Continuous Wavelet Transform, Filtering, and Rate Calculation. These steps are detailed below. Finally, the correlation between signals retrieved from thermal images and those obtained by the ground truth contact based physiological sensors were also calculated.

1) Blood vessels detection: We experimented different regions in the thermal faces for extracting the heart rate. First, we detected the heart rate from thermal images by tracking superficial blood vessels on the face [16]. To segment the blood vessels from the the forehead region, we used several well-known edge detection methods, including Canny, Prewitt, Roberts, and Sobel. The Canny edge detection method proved to be the most effective in our experiments. The forehead was chosen in particular as it provides a semi-flat surface for cleaner detection of blood vessels. Following the segmentation, we were able to generate our vascular map. Furthermore, edge detection may miss the core of the vein, where the effect of heat transfer caused by blood circulation is most prominent. As a result, we enlarged the edges by one pixel in each direction to ensure that we extracted the heat coming from the center of each vein near the skin's surface.

2) Inner canthus detection: We experimented the eyes corners regions. The eye corner (inner canthus) represent the warmest region through the recording [7]. In order to locate the inner canthus, we used image binarization in the region of the eyes with a threshold equal to 1:

$$Threshold = Maximum - (Maximum - Median)/2 \quad (1)$$

3) Signal Extraction Pipeline:

i) **Differencing:** We calculate the differences between adjacent elements of the signal $S(t)$ to produce the transformed signal $\hat{S}(t)$

$$\hat{S}(t) = S(t) - S(t-1) \quad (2)$$

ii) **Normalization:** The signal's amplitude was normalized using μ and σ as the mean and standard deviation of S_i respectively. The transformed signal $S(t)$ has mean $\mu = 0$ and standard deviation $\sigma = 1$.

$$\hat{S}(t) = \frac{S(t) - \mu}{\sigma} \quad (3)$$

iii) **Downsampling:** Downsampling was performed on the thermal signal to reduce the signal rate to 8 Hz from 100 Hz for the original signal. This matched the frequency of the physiological signal, which itself was downsampled to 8 Hz from 2048 Hz as part of the pre-processing procedure in order to increase the computational efficiency without noticeable degradation of information.

iv) **Continuous Wavelet Transform:** The Mexican Hat a.k.a. the Ricker Wavelet was used as the mother wavelet $\psi(t)$. Equation 4 describes the Ricker Wavelet in which σ represents the standard deviation and t represents time. Equation 5 describes the continuous wavelet transform in which S is the input signal function, t is time, a is the scale value and b is the translation value. The scale which best represents the physiological signal component of the recordings was selected based on a separate validation set. Lower scales of a wavelet transform are more likely to contain noise, while higher scales are more likely contain metabolic contributions. Considering that a normal heart rate has a higher frequency than normal breathing rate, and in order to avoid excessive 'smoothing', we utilized a lower scaling for heart rate and a higher scaling for breathing rate.

$$\psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{1/4}} \left(1 - \left(\frac{t}{\sigma}\right)^2\right) e^{-\frac{t^2}{2x^2}} \quad (4)$$

$$S_w(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} S(t)\psi\left(\frac{t-a}{b}\right) dt \quad (5)$$

v) **Filtering:** Next, we apply an Elliptic filter to isolate the breathing rate. An Elliptic bandpass was selected for its sharper transition between filtered and unfiltered frequencies. On the other hand, for the heart rate, we used the butterworth bandpass filter [36].

vi) **Rate Calculation:** Next, the number of peaks in the wave are counted. However, the smaller peaks found in some waveforms are not necessarily consistent with the breathing function and are more closely attributed to noise. This issue is addressed with a constraint known as peak prominence which is introduced when selecting the peaks. Peak prominence measures the height of a peak relative to its nearby surrounding peaks. As the

signal levels vary across our subjects as well as between the sensor and thermal signals, utilizing this additional step is suitable for our dataset.

a) **Respiration:** For respiration, the prominence parameter is set to a certain threshold of the range of the signal in order to filter out noise relative to the signal level. The respiration rates are then calculated by counting the resulting peaks and dividing by the length of the recording. The distribution of rates across recordings can be found in Fig. 2a, which shows the majority of recorded subjects falling within the normal range of 12 to 20 breaths per minute [39].

b) **Heart Rate:** Finally, all of the extracted signals from thermal recordings were found to fall within the normal range of 60 to 100 beats per minute, with the exception of one outlier, as shown in Fig. 2b [17]. It may also be noted here that the thermal recordings seem to outperform the ground truth when determining the number of recordings that fall within normal range, with about 90% of the contact based physiological signal recordings falling between 60 and 100 beats per minute. While further investigation would be needed, this may be due to noise caused by subject movement, an issue which is more likely to occur with contact-based sensors.

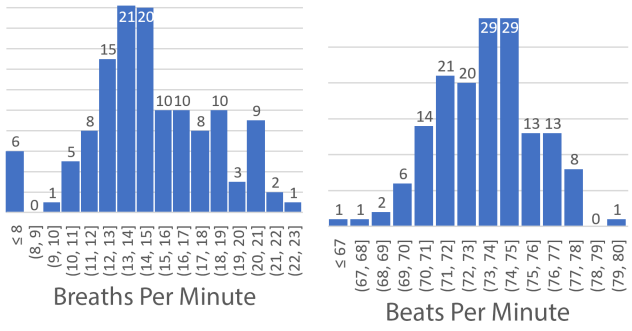
TABLE I: Differences in Breathing Rates Across Observed and Ground Truth

BPM Difference	Respiration
<0.5	51
<1	77
<2	93
Total Recordings	128

Cross correlation was used to find the maximum correlation between shifted copies of the sensor and thermal signals for the respiration rate. By utilizing cross correlation, we aimed to perfectly align the signals at sub-second accuracy, as shown in Fig. 3a and Fig. 3b.

4) Skin Temperature Signal: According to studies in the literature, changes in core and total skin temperature occur extremely gradually over a period of minutes or even hours in steady, comfortable environmental conditions [38]. As a result, we used multiple regression models to determine the long-term trend of the thermal signal retrieved from the thermal faces and to evaluate the performance of the NC approach to extract skin temperature compared to the sensor-based skin temperature. We used the whole face region to extract the NC skin temperature and to report our results using Linear regression, Polynomial regression, Lasso regression, Ridge regression, and ElasticNet regression.

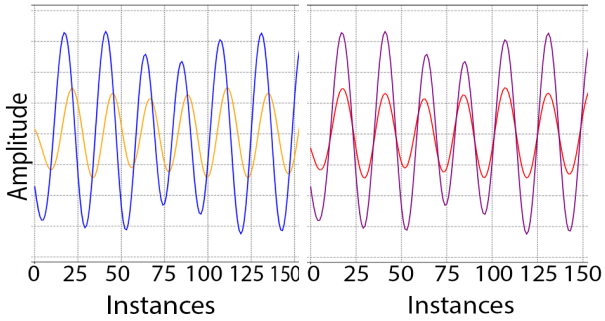
5) Final Non-Contact Physiological Feature vectors: Following the extraction of the NC based physiological signals, we formed their final feature vectors using the detected beats per minute for the heart rate and breaths per minute for the respiration rate. Moreover, we generated statistical



(a) Distribution of breathing rates

(b) Distribution of heart rates

Fig. 2: Histograms showing the rates per minute of Respiration and Heart Rate, showing the number of subjects (vertically) that fall into BPM bins (horizontally)



(a) Recording 1-04-b before cross correlation adjustment

(b) Recording 1-04-b after cross correlation adjustment

Fig. 3: Alignment of signals before and after cross correlation

features from the extracted signals. The features include a 10 bin histogram that describes the signal distribution, the minimum, the maximum, the mean, the variance, the skewness, the kurtosis, the inter-quartile range, the standard error of the mean, and the median.

D. Classification

The final set of experiments we present in this paper utilizes the contact based physiological signals, the thermal features, and the extracted NC based physiological features to train different classifiers in order to detect the subjects' enervation. In addition to running classification for these three sets of features, we generated a larger feature vector per recording by merging features from the thermal modality with the NC based physiological signals before training our classifiers in order to provide a meaningful comparison. In our experiments, we defined our classification labels based on the Time hypothesis (hereafter referred to as the Time label), in addition to the KSS based labels (hereafter referred to as the KSS label), which are derived from the KSS survey results.

The Time label was built on the presumption that participants were energized within an hour of waking up and enervated in the evening after a long busy day, as discussed earlier in Section III. While, for the KSS labels, a KSS

score was obtained from the subjects for each session, with values ranging from one to nine; here, one indicated that the subject was the most energized and nine indicated the most enervation. We converted this range into a binary classification problem, with one to five being energized and six to nine being enervated.

In this paper, we are using a total of 159 recordings which consist of 33 baseline recordings, 64 morning recordings, and 62 evening recordings. A few recordings were missing due to errors in data collections.

In our experiments, we explored several supervised machine learning classifiers; however, we settled on Random Forest Classifier (RFC), and Extreme Gradient Boosted Machine (XGB) as these two classifiers performed well in literature [25]. To establish the baseline metrics, a baseline classification utilizing random guessing was used. This Baseline Classifier (BC) serves as a simple baseline against which other more complex classifiers can be compared. Leave-One-Subject-Out Cross Validation was used to evaluate performance; this allowed for the training set to exclude one subject's recording set at a time, with that subject's recording then used for testing for a given fold.

V. Experimental Results

Our results for this paper are categorized in two sections. The first section presents the results of our evaluation of the NC physiological features extracted from the thermal modality, while the second section presents the results of the classification models using the different sets of features, described earlier, to detect enervation.

A. Evaluation of Non-Contact Physiological Features

1) Non-Contact Respiration and Heart Rate Scoring Method: To understand the performance of our system, we scored our results with a cumulative threshold method. This involved determining the signals' rates, as described in section IV, which were detected for both the contact-based physiological and NC based physiological signals of the same recording. Next, we took the total difference in the number of peaks between the two modalities relative to the length of the recording to devise a score.

(A) Respiration Rate: Based on the difference between the NC signal and the contact-based signal, we determine whether the rates from both signals represent a match, by falling within a defined threshold value, or not, by exceeding the threshold value. The accuracy is then calculated as the total number of matches divided by the total number of recordings. The complete set of results are shown in Fig. 4. Using this scoring method, the figure shows we found an overall accuracy of 80% for the respiration and for the NC signal for the Heart Rate generated from the median inner canthus at the corresponding threshold.

Additionally, the number of recordings within certain breath-per-minute differences of the contact based ground truth are found in Table I.

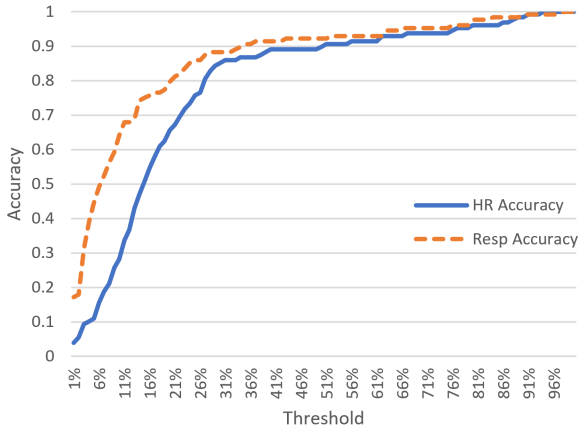


Fig. 4: NC Heart Rate and NC Respiration Accuracy as a Function of Threshold

2) Skin Temperature: In order to evaluate whether the extracted NC skin temperature matches the contact based skin temperature we used regression for each recording's data separately, with 80% as training data and 20% as testing data.

Table III presents the average mean squared error for the evaluation of different models using the following thermal features taken across the frame: the average, the maximum, and the 10% highest temperature. The mean squared error provides a better understanding of the correlation between the two signals. Thus, a lower error indicates higher correlation. The linear and polynomial models have the lowest error of 0.54 and 0.44 degrees Fahrenheit respectively using the 10% highest temperature feature, which shows high correlation between the contact and NC physiological signals.

B. Classification Of Enervation State

For evaluation of enervation detection, we report the average overall accuracy and the mean recall using each of the three individual contact based physiological signals, each of the three NC based physiological signals, the thermal features, and different combinations of the contact as well as the NC signals.

Based on our preliminary results, the heart rate signal extracted from the inner canthus showed better performance compared to the NC extracted heart rate signal from the other thermal regions, described earlier. Thus, for our enervation classification, we will utilize the results of the heart rate signal extracted from the inner canthus.

Regarding the thermal features, the forehead region demonstrated better performance compared to the other regions in literature [1]. Therefore, this region is utilized to represent the thermal features for our enervation classification.

Table II describes the overall accuracy and the mean recall for the different classification schemes. It shows a comparison of NC based features against the contact based features as follows: First, the performance of each set of physiological features were individually assessed. As presented in the

TABLE II: Comparison of Contact and NC-based Classifications Of Enervation State, H: heart rate signal, R: Respiration rate signal, S: Skin temperature signal, F: Thermal forehead region, TL: Time Label, KL: KSS Label, GB: Gradient Boosted, RF: Random Forest, BC: Baseline Classifier. The highest accuracy in each column is presented in bold.

			Non-Contact Based											**Contact Based**						
			H	**R**	**S**	**F**	**HR**	**HS**	**SR**	**HSR**	**FH**	**FR**	**FS**	**FHR**	**H**	**R**	**S**	**HR**	**HS**	**SR**	**HSR**
TL	**GB**	**Accuracy**	50.8	**49.2**	57.9	61.1	53.2	**63.5**	**59.5**	**58.7**	63.5	**66.7**	60.3	**63.5**	50.0	36.5	44.5	48.4	50.8	42.1	50.0
		Recall	50.7	49.2	57.9	61.1	53.1	63.5	59.6	58.7	63.4	66.6	60.3	63.4	50.0	36.5	44.5	48.4	50.7	41.9	50.0
	RF	**Accuracy**	**59.5**	42.1	**58.7**	**65.9**	**57.9**	61.1	56.3	54.0	**69.8**	65.9	**63.5**	62.7	55.6	44.4	42.9	51.6	51.6	46.8	47.6
		Recall	59.5	42.0	58.7	65.8	57.8	61.0	56.3	53.9	69.8	65.9	63.5	62.6	55.5	44.3	42.8	51.5	51.5	46.7	47.6
	BC	**Accuracy**	50.8	42.1	46.0	49.2	56.3	45.2	50.8	51.0	47.6	49.2	51.6	50.0	**57.1**	46.8	50.0	53.2	50.8	43.7	44.4
		Recall	50.9	42.1	46.0	49.1	56.4	45.1	50.9	50.7	47.6	49.2	51.6	50.0	57.2	46.8	50.0	53.3	50.7	43.5	44.3
KL	**GB**	**Accuracy**	42.5	49.2	49.2	53.3	45.0	47.5	45.0	52.5	45.8	52.5	53.3	49.2	52.5	**53.3**	45.8	53.3	55.8	47.5	52.5
		Recall	40.7	45.9	47.0	50.9	42.3	44.4	40.9	49.0	43.3	49.9	51.1	46.1	50.7	52.6	43.0	51.4	54.4	45.6	49.9
	RF	**Accuracy**	43.3	51.7	44.2	50.8	47.5	45.0	45.8	45.8	55.0	59.2	50.8	52.5	54.2	50.0	**57.5**	**54.2**	**55.0**	**49.2**	53.3
		Recall	39.7	47.7	39.9	47.6	43.6	41.1	41.9	41.3	50.6	55.6	46.1	48.7	51.2	46.6	55.0	51.3	52.6	46.7	50.9
	BC	**Accuracy**	46.7	43.3	50.8	47.5	51.7	46.7	53.3	48.3	48.3	50.8	46.7	47.5	52.5	52.5	52.5	45.8	55.0	45.0	**55.8**
		Recall	47.4	43.14	50.1	48.4	51.1	47.7	52.6	49.1	49.1	51.0	46.3	47.9	53.9	52.1	53.6	46.7	54.9	44.3	55.9

TABLE III: Per Recording Average Mean Squared Error Evaluation Metric Using Regression Models

Features	**Average**	**Maximum**	**10% maximum**
Linear Model	0.8105	0.6464	0.5469
Polynomial Model	0.7316	0.5251	0.4498
Lasso Model	0.8089	0.6349	0.5403
Ridge Model	0.8435	0.6383	0.5553
ElasticNet Model	0.9507	0.8922	0.9629

table, the NC based modalities outperformed the contact based modalities using the Time label. Classification using the thermal features (presented as 'F' in the table) attained 65.87% accuracy using Random Forest, outperforming the other individual modalities. Regarding the KSS label, the skin temperature using the contact based features outperform the other individual modalities with an accuracy of 57.5%.

Moreover, we evaluated the performance of merging the different physiological signals in addition to merging the thermal features with the NC based physiological signals. The aforementioned table also shows that the NC based combinations outperform the contact based method for both the KSS and Time labels. The combination of thermal features and NC heart rate signal (presented as 'FH' in the table) in particular attained an accuracy of approximately 70% using the Random Forest classifier with the Time label; this was also an improvement compared to the individual NC signals. In assessing classification results, we found that NC respiration did not perform well as an individual modality classifying enervation, but improved other results when merged with other signals. Its poorer performance may be explained by the fact that the subjects' breathing was irregular during recordings where the subject is speaking. The KSS label was not very promising in term of enervation classification compared to the Time label, as the greatest accuracy attained was approximately 60%, using the combination of the forehead thermal features with the NC respiration rate signal (presented as 'FR' in the table). As for classifying enervation with contact-based sensors, performance was comparable to the baseline classifier, which may be caused by noise due to the subjects movement, an inherent problem to contact-based sensors. This motion may have distorted patterns related to enervation detection.

Considering the complete set of results, we can see that the performance of the individual signals is enhanced through the merging of additional features, and in particular this was the case for the NC features. Accordingly, the NC features provided promising results in terms of enervation classification, which is practical when being applied in a vehicle.

VI. Conclusion

In this paper, we use a multimodal dataset for circadian rhythm detection, as well as propose a pipeline utilizing waveform transformations and bandpass filtering as part of a framework able to extract heart rate, respiration rate, and skin temperature from thermal images to provide an NC based alternative. In addition, we classified enervation using contact and NC approaches, and demonstrated the efficacy of the NC approach. This provides the opportunity to move towards an implementable technology in autonomous vehicles that does not rely on uncomfortable, restrictive contact-based sensors.

Our research focused on said comparison by investigating and comparing in-depth results of both contact and NC modalities. We found that thermal images processed through our pipeline generated physiological signals with comparable output to the ground truth contact-based sensors. More specifically, the NC respiration signal was found to match the contact-based signal with a high accuracy, while the NC heart rate signal extracted from the median inner canthus was found to match the contact-based signal with high accuracy. Finally, the skin temperature signal showed a high correlation between the retrieved signal with the contact physiological skin temperature with an average mean squared error of 0.44 and 0.54 using polynomial and linear models respectively.

In addition, as part of a second stage of analysis, we found that classifying the subjects' enervation states across the Time label using these NC physiological signals demonstrated promising results, reaching approximately 70% accuracy. It achieved a marked improvement in performance when compared against the KSS label's self-reported sleepiness levels. Among the modalities used in classification, the forehead thermal region performed best, and the NC respiration performed worst. While further investigation would be needed, this may be explained by the inherent instability of the respiration signal when the subject is speaking, compared

to the inherent stability of the forehead-based signal. This work contributes to an improved knowledge of the status of a vehicle's occupants in order to enhance their comfort and well-being.

VII. Acknowledgments

This material is based in part upon work supported by the Ford Motor Company. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Ford Motor Company or any other Ford entity.

References

[1] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo. Detecting deceptive behavior via integration of discriminative features from multiple modalities. *IEEE Transactions on Information Forensics and Security*, 12(5):1042–1055, 2017.

[2] M. Akashi, R. Sogawa, R. Matsumura, A. Nishida, R. Nakamura, I. T. Tokuda, and K. Node. A detection method for latent circadian rhythm sleep-wake disorder. *EBioMedicine*, 62:103080, 2020.

[3] T. Åkerstedt and M. Gillberg. Subjective and objective sleepiness in the active individual. *International journal of neuroscience*, 52(1-2):29–37, 1990.

[4] J. Arendt, A. Borbely, C. Franey, and J. Wright. The effects of chronic, small doses of melatonin given in the late afternoon on fatigue in man: a preliminary study. *Neuroscience letters*, 45(3):317–321, 1984.

[5] S. L. Bennett, R. Goubran, and F. Knoefel. Comparison of motion-based analysis to thermal-based analysis of thermal video in the extraction of respiration patterns. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3835–3839. IEEE, 2017.

[6] M. A. Bonmati-Carrion, B. Middleton, V. Revell, D. J. Skene, M. Rol, and J. A. Madrid. Circadian phase asessment by ambulatory monitoring in humans: Correlation with dim light melatonin onset. *Chronobiology international*, 31(1):37–51, 2014.

[7] S. Budzan and R. Wyżgolik. Face and eyes localization algorithm in thermal images for temperature measurement of the inner canthus of the eyes. *Infrared Physics & Technology*, 60:225–234, 2013.

[8] P. Cheng, O. Walch, Y. Huang, C. Mayer, C. Sagong, A. Cuamatzi Castelan, H. J. Burgess, T. Roth, D. B. Forger, and C. L. Drake. Predicting circadian misalignment with wearable technology: validation of wrist-worn actigraphy and photometry in night shift workers. *Sleep*, 44(2):zsaa180, 2021.

[9] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque. Sensor applications and physiological features in drivers' drowsiness detection: A review. *IEEE sensors Journal*, 18(8):3055–3067, 2018.

[10] M. Egger, M. Ley, and S. Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.

[11] J. Fei and I. Pavlidis. Thermistor at a distance: unobtrusive measurement of breathing. *IEEE Transactions on Biomedical Engineering*, 57(4):988–998, 2009.

[12] E. E. Flynn-Evans, H. Tabandeh, D. J. Skene, and S. W. Lockley. Circadian rhythm disorders and melatonin production in 127 blind women with and without light perception. *Journal of Biological Rhythms*, 29(3):215–224, 2014. PMID: 24916394.

[13] R. R. Freedman, D. Norton, S. Woodward, and G. Cornélissen. Core body temperature and circadian rhythm of hot flashes in menopausal women. *The Journal of Clinical Endocrinology & Metabolism*, 80(8):2354–2358, 08 1995.

[14] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda, et al. Heart rate variability-based driver drowsiness detection and its validation with eeg. *IEEE Transactions on Biomedical Engineering*, 66(6):1769–1778, 2018.

[15] T. Gault and A. Farag. A fully automatic method to extract the heart rate from thermal video. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 336–341, 2013.

[16] T. R. Gault, N. Blumenthal, A. A. Farag, and T. Starr. Extraction of the superficial facial vasculature, vital signs waveforms and rates using thermal imaging. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 1–8. IEEE, 2010.

[17] M. Gertsch. The normal electrocardiogram and its (normal) variants. *The ECG Manual: An Evidence-Based Approach*, pages 17–36, 2009.

[18] R. Hartley and A. Zisserman. Multiple view geometry in computer vision (cambridge university, 2003). *C1 C3*, 2, 2013.

[19] C. Hessler, M. Abouelenien, and M. Burzo. A non-contact method for extracting heart and respiration rates. In *2020 17th Conference on Computer and Robot Vision (CRV)*, pages 1–8. IEEE, 2020.

[20] P. Jakkaew and T. Onoye. Non-contact respiration monitoring and body movements detection for sleep using thermal imaging. *Sensors*, 20(21):6307, 2020.

[21] S. I. Kaduk, A. P. Roberts, and N. A. Stanton. The circadian effect on psychophysiological driver state monitoring. *Theoretical Issues in Ergonomics Science*, pages 1–25, 2020.

[22] C. F. Kerry and J. Karsten. Gauging investment in self-driving cars. *Brookings Institution, October*, 16, 2017.

[23] V. Kolodyazhniy, J. Späti, S. Frey, T. Götz, A. Wirz-Justice, K. Kräuchi, C. Cajochen, and F. H. Wilhelm. An improved method for estimating human circadian phase derived from multichannel ambulatory monitoring and artificial neural networks. *Chronobiology International*, 29(8):1078–1097, 2012.

[24] H. Masuda, S. Okada, N. Shiozawa, M. Makikawa, and D. Goto. The estimation of circadian rhythm using smart wear. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4239–4242. IEEE, 2020.

[25] M. Papakostas, K. Riani, A. B. Gasiorowski, Y. Sun, M. Abouelenien, R. Mihalcea, and M. Burzo. Understanding driving distractions: A multimodal analysis on distraction characterization. In *26th International Conference on Intelligent User Interfaces*, pages 377–386, 2021.

[26] K. Riani, M. Papakostas, H. Kokash, M. Abouelenien, M. Burzo, and R. Mihalcea. Towards detecting levels of alertness in drivers using multiple modalities. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–9, 2020.

[27] K. Riani, S. Sharak, K. Das, M. Abouelenien, M. Burzo, R. Mihalcea, J. Elson, C. Maranville, K. Prakah-Asante, and W. Manzoor. Towards classifying human circadian rhythm using multiple modalities. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. IEEE, 2021.

[28] A. Rivera-Coll, X. Fuentes-Arderiu, and A. Díez-Noguera. Circadian rhythms of serum concentrations of 12 enzymes of clinical interest. *Chronobiology International*, 10(3):190–200, 1993.

[29] J. Shi et al. Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE, 1994.

[30] S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc. Gpu-based video feature tracking and matching. In *EDGE, workshop on edge computing using new commodity architectures*, volume 278, page 4321, 2006.

[31] J. E. Stone, A. J. Phillips, S. Ftouni, M. Magee, M. Howard, S. W. Lockley, T. L. Sletten, C. Anderson, S. M. Rajaratnam, and S. Postnova. Generalizability of a neural network model for circadian phase prediction in real-world conditions. *Scientific reports*, 9(1):1–17, 2019.

[32] N. Sun and I. Pavlidis. Counting heartbeats at a distance. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 228–231. IEEE, 2006.

[33] G. Tanda. The use of infrared thermography to detect the skin temperature response to physical activity. In *Journal of Physics: Conference Series*, volume 655, page 012062. IOP Publishing, 2015.

[34] M. Tashakori, A. Nahvi, and S. Ebrahimian Hadi Kiashari. Driver drowsiness detection using facial thermal imaging in a driving simulator. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of engineering in medicine*, 236(1):43–55, 2022.

[35] B. C. Tefft. Acute sleep deprivation and culpable motor vehicle crash involvement. *Sleep*, 41(10), 09 2018. zsy144.

[36] P. van Gent, H. Farah, N. Nes, and B. van Arem. Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data. In *Proceedings of the 6th HUMANIST Conference*, pages 173–178, 2018.

[37] M. H. Vitaterna, J. S. Takahashi, and F. W. Turek. Overview of circadian rhythms. *Alcohol Research & Health*, 25(2):85, 2001.

[38] D. Wang, H. Zhang, E. Arens, and C. Huizenga. Observations of upper-extremity skin temperature and corresponding overall-body thermal sensations and comfort. *Building and Environment*, 42(12):3933–3943, 2007.

[39] G. Yuan, N. A. Drost, and R. A. McIvor. Respiratory rate and breathing pattern. *McMaster Univ. Med. J*, 10(1):23–25, 2013.