2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A novel Hot-Flash classification algorithm via multi-sensor features integration

Andreas Tsiartas, Fiona C. Baker, David Smith, Massimiliano de Zambotti

Abstract— **We aim to evaluate the feasibility and performance of a novel hot flash (HF) classification algorithm based on multi-sensor features integration using commercial wearable sensors. First, we processed feature sets from wrist-based multi-sensor data (photoplethysmography, motion, temperature, skin conductance and). Then, we classified (Decision Tree) physiological-recorded HFs (N=27) recorded from three menopause women, and we assessed the algorithm performance against gold-standard HF expert evaluation. The results indicated that while skin conductance features alone explain most of the variance (~65%) in HF classification, the multi-sensor approach achieved above 90% sensitivity at 95.6% specificity in HF classification and showed advantages under conditions of signal corruption and different biobehavioral states (sleep vs wake). The proposed new multi-sensor approach showed being promising in HF classification using common commercially-available wearable sensors and target locations.**

Clinical Relevance— **The development of "user-centered" accurate, automatic detection systems for HFs can advance the measurement and treatment of HFs.**

I. INTRODUCTION

Hot flashes (HFs) are a thermoregulatory phenomenon that are a hallmark of the menopause transition. HFs are characterized by peripheral vasodilation and sweating, lasting 1-5 minutes, that can occur hourly or daily [1]. HFs are the most common and disruptive menopausal symptom, affecting upward of 80% of women in midlife, and potentially persisting for several years [2].

The current gold-standard method for measuring HFs is the expert evaluation of sudden increases (2 uS/30s) in sternal skin conductance (SC) recorded via laboratory or ambulatory research-grade devices [3]. The development of automatic algorithms for HFs classification is on the rise, and current algorithms, including our own prior HF detection algorithm [4], are mainly based on sternum SC signal processing (e.g., using fixed SC threshold, pattern recognition techniques, neural networks, template matching). Of note, a multitude of other physiological changes (e.g., increases in heart rate, skin temperature) accompany the HF manifestation [4, 5], although there is variability in these changes based on factors like motion and biobehavioral state (e.g. wake vs. sleep (see [5]).

No current solution exists within the consumer space for measuring HFs. In the current work, we introduce and evaluate the feasibility and performance of a novel multi-sensor approach to HF detection based on multi-feature integration from consumer-grade SC, temperature (T), motion (M) and photoplethysmography (PPG) sensors placed on the wrist. The wrist was chosen as a promising target location that could enable data collection via consumer smart wearable technology. Expert evaluation of sternum SC fluctuation was used as the gold standard reference for comparison. In addition, we evaluated sensor performance for HF detection under two conditions: (1) sleep vs wake (2) sensor loss of contact or faulty sensors.

II. METHOD

A. Sample

Three women (Age, mean ± SD: 55.6 ± 0.6 y) who reported having daily HFs participated in a ~12h lab-based study, that encompassed an overnight. A total of 27 physiological HFs were recorded from the women. The study was reviewed and approved by Advarra Institutional Review Board (Protocol number: Pro00040686), and participants provided written, informed consent.

Women were free from major mental and medical conditions, had undergone natural menopause, and none of them was currently taking hormone therapy. None of them had breathing or leg-movement related sleep disorders, confirmed by PSG.

B. Procedure

All recordings took place at the SRI International Human Sleep Research Laboratory. Standard PSG data collection (including electroencephalography, electromyography and electrooculography) was performed using Compumedics Grael 4K PSG:EEG (Abbotsford, Victoria, Australia), and sleep was scored according the American Academy of Sleep Medicine (AASM) guidelines [6].

Physiological HFs were recorded and scored (2 uS/30s rises in SC) by experienced scorers, according to gold standard methods: sternal SC (64 Hz) was collected via two 1.5 cm-diameter Ag/AgCl electrodes filled with 0.05 M potassium chloride Velvachol/glycol gel placed on either side of the sternum (about 4 cm apart; a 0.5-V constant voltage circuit was maintained between them) using a BioDerm SC Meter (model 2701; UFI, Morro Bay, CA) [see 5].

Signals from a customized array of consumer-grade commercially available sensors (PPG: S/F SEN-11574 - 512

Andreas Tsiartas (PhD). Fiona Baker (PhD), David Smith, and Massimiliano de Zambotti (PhD) were with the SRI International, CA, USA. (correspondence e-mail: andreas.tsiartas@sri.com, author: Andreas Tsiartas)

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

Hz; SC sensor: Grove 101020052 - 64 Hz; 3-axis motion sensor: NXP-FXOS8700 - 1024 Hz; T sensor: TI-TMP36GT9Z – 16 Hz) were collected from each participant's wrist (M and PPG sensors on the dorsal wrist; SC [also called Galvanic skin response] and T on the anterior wrist) and integrated with the Compumedics recording system, using a multi-channel output card (40-Ch Digital-to-Analog Converter: A/D-AD5370). A sample of physiological data acquisition is depicted in Fig. 1.

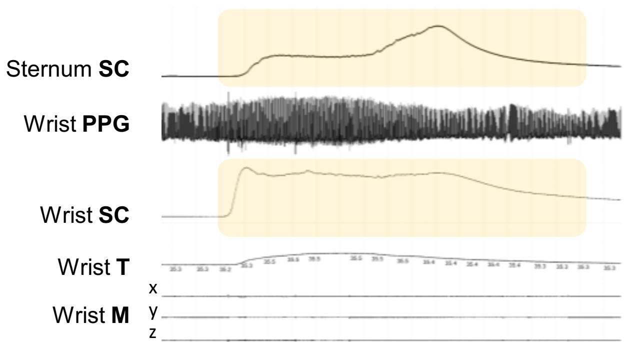

Fig. 1 Co-registration of gold standard sternum skin conductance (SC) and consumer-grade commercially available multi-sensor signals from the wrist across a hot flash (HF) event (highlighted in the graph; HF duration: ~8 min) in one participant. T, temperature; PPG, photoplethysmography; M, motion.

The data collection started ~3h before bedtime and continued overnight, until the morning awakening. Women slept in sound-attenuated and temperature-controlled bedrooms.

C. *Processing of multi-features from consumer-grade sensors*

We extracted features from all four sensors (SC, PPG, T, and M sensors). Every 15 seconds, we computed each feature using a windowing approach. We used left and right windows for each feature computation. We obtained the distinct feature sets for SC (SC and SC+), T, PPG, and M.

Next, we describe the processed feature sets from multi-sensor data; SC, skin conductance; T, temperature; PPG, photoplethysmography; M, motion:

- **SC feature set** uses the HF onset output of a previously developed HF prediction algorithm [4], as a feature. We designate the HF onset as the ±2 minutes around the HF predicted onset.
- **SC+ feature set** includes the SC set and the differential Area Under the Curve (AUC) of the SC signal. To compute the differential, we take the AUC difference between the last 250 s and the current window (±30 s). In addition, we double the derivative described in step 8 of our prior SC algorithm [4]. The SC+ feature set aims to represent both slow and fast rising SC responses using the AUC and the SC feature set, respectively.
- **T feature set** computes the participant's temperature average differential between the prior and following 500 s. The feature aims to capture temperature changes before and after a HF event.
- **PPG feature set** uses a FFT-based heart rate (HR) estimation. The HR estimation is averaged in two regions: 120 s before and after the window. The differential of the HR change is used as a feature. The PPG feature aims to capture HR changes before and after a HF event.
- **M feature set** captures movements of the subject in the x, y, and z dimensions. We process each dimension separately and extract the absolute maximum displacement (AMD) (window ±30 s) for each dimension. For the x dimension, we use the raw AMD. For y and z, we use the AMD differential between y, z, and x as features.

We time-aligned the features with the HF expert annotations for prediction and evaluation (±90 s matching window). Since all the features are processed independently, the system enables sensor specific feature selection. The selected features were fed into a Decision Tree classifier, which makes a decision every 15 s, whether or not the current frame is a HF by using multi-sensor information. Finally, using the decision output, we extracted the HF regions for each subject.

III. ANALYSES

The following analyses were conducted:

1. *Evaluated SC features vs multi-sensor features in the HF classification performance.*

First, we compared HF classification accuracy for the SC+ set vs the SC set. Then, we augmented the SC+ set with T, PPG, and M sets.

2. *Evaluated the HF classification performance as a function of whether the HF onset occurred during wake or sleep.*

We repeated the analyses of (1) with the same set of features (and the same system) but, in this case, we compared the sleep and awake regions as scored from PSG. We compared the impact of the sensors on the two conditions.

3. *Evaluated the HF classification robustness in a simulated noise environment.*

We repeated the experiments of (1) with the same set of features and system but, in this case, we simulated and corrupted the signals with sensor contact loss. To simulate sensor-contact loss (partial contact), we randomly selected 15% of a participant's session region and assigned the lowest value of the signal, for each sensor independently.

In all the analyses, we randomly split our data in two sets, 80% of data for training the Decision Tree and 20% for evaluating the system. We repeated this process using a 5-fold, cross-validation setup until all data were used for testing. For the decision tree training, we used maximum depth 6 (6 decisions from root to leaf) and 5 minimum samples per leaf (a decision applies to 5 or more samples in the data; if the decision applies to less than 5 samples, we discard the decision). To ensure similar specificity (96.5±1%) across analyses, we oversampled the HF class to 1:2 ratio between the HF and non-HF regions. HF performance was evaluated in terms of system sensitivity (percent of true positives) and specificity (percent of true negatives) in HF detection compared to gold standard sternum SC expert evaluation.

We computed the impact (contribution) of each sensor using the Shapley values method [7], assigning the optimal impact to each sensor given the consistency and additivity assumptions. For equal class representation, we ran the analyses by sampling to a 1:1 ratio between the two classes.

IV. Results

A. SC features vs multi-sensor features

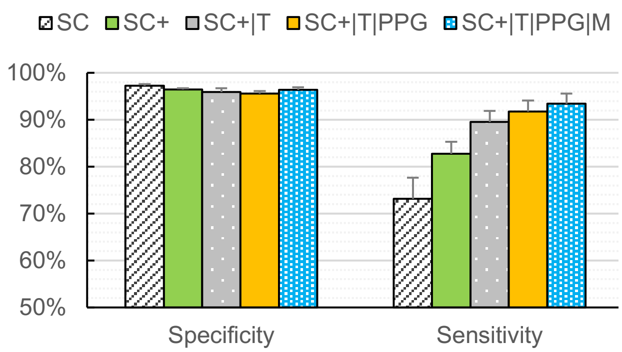

Fig. 2 Hot flash classification performance as a function of the feature set or feature sets combination. Vertical bars represent mean and standard deviation. SC, skin conductance; T, temperature; PPG, photoplethysmography; M, motion.

At 96.5% specificity, the SC+ set showed better HF classification performance than the SC set (+10.7% in sensitivity). The addition of the T, PPG, and M sets to the SC+ set resulted in further improvements (see Fig. 2).

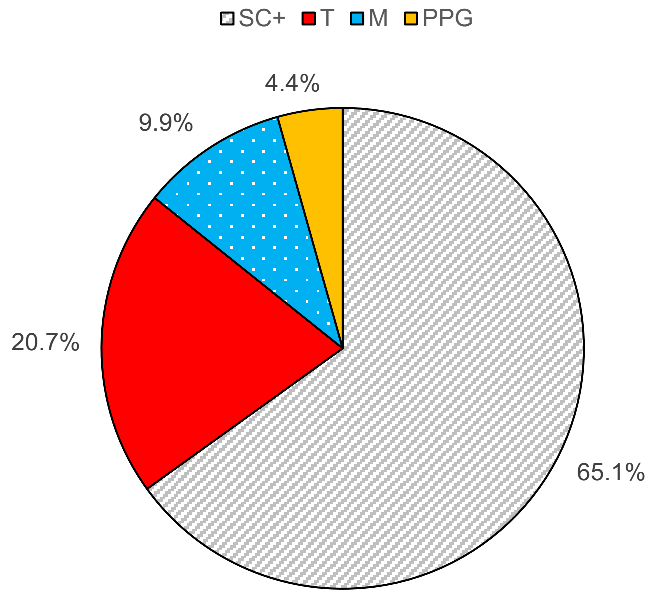

Fig. 3 Sensor contribution to the "SC+|T|PPG|M" system. SC, skin conductance; T, temperature; PPG, photoplethysmography; M, motion.

The importance of the multi-sensor system is shown in Shapley values of each sensor for the best system (SC+|T|PPG|M) (see Fig. 3). While the SC signal accounted for the most variance in HF classification (~65%), using additional non-SC features further enhanced the classification performance.

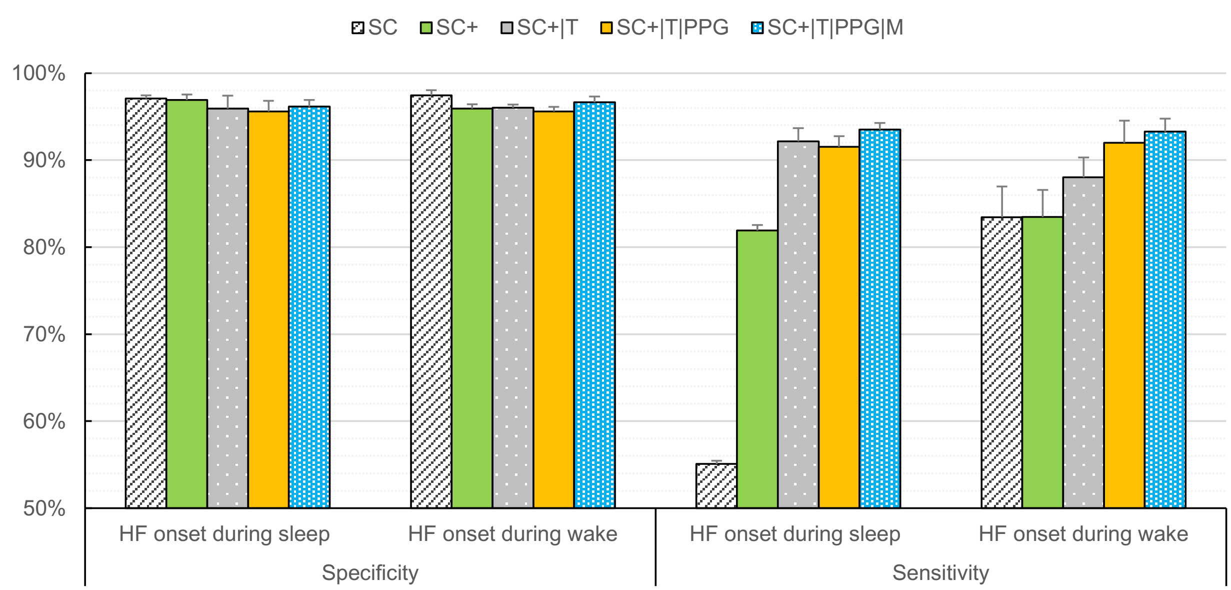

Fig. 4 System performance for hot flashes (HFs) onsets occurring during sleep vs wake. Vertical bars represent mean and standard deviation. SC, skin conductance; T, temperature; PPG, photoplethysmography; M, motion.

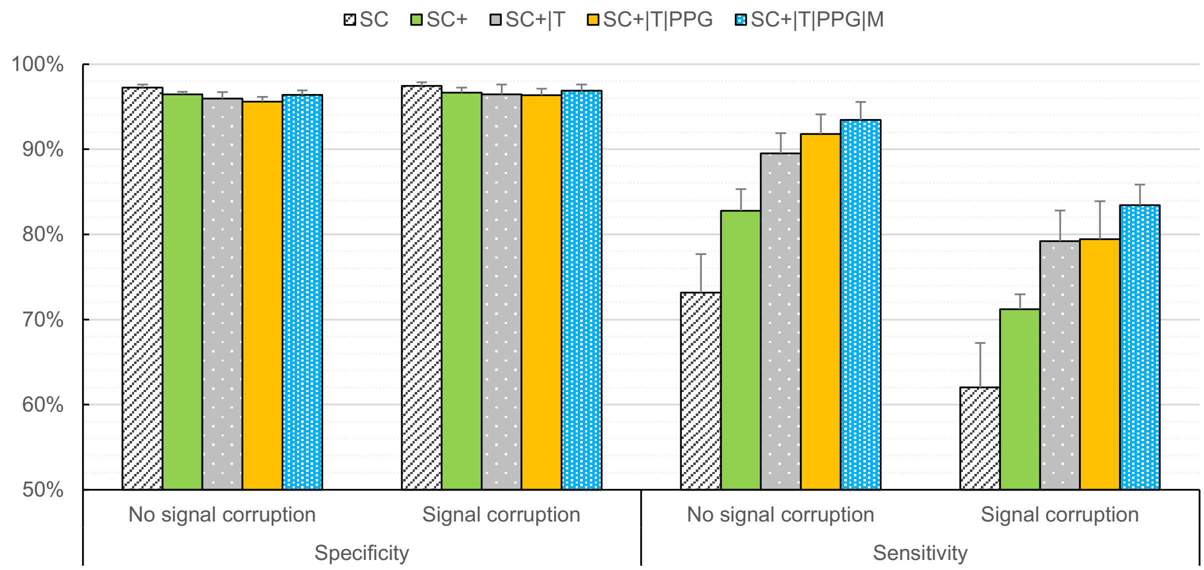

Fig. 5 Hot flash (HF) classification performance as a function of the feature set or feature sets combination, in conditions of reliable and corrupted signals. Vertical bars represent mean and standard deviation. SC, skin conductance; T, temperature; PPG, photoplethysmography; M, motion.

B. *HFs occurring during sleep vs wake*

We observed a greater contribution from the non-SC features in the HF classification performance for HFs with onsets occurring during sleep vs wake (see Fig. 4).

From the Shapley values, it is noticeable that the features contribution in the HF classification was greater for the PPG, SC, and M sets, while it was less for the T set, when comparing HFs with onsets occurring during wake vs sleep (Fig. 6).

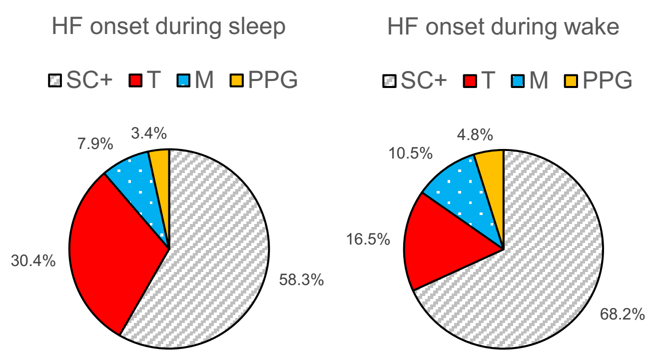

Fig. 6 Sensor contribution to the "SC+|T|PPG|M" system for hot flash (HF) onset occurring during sleep vs wake. SC, skin conductance; T, temperature; PPG, photoplethysmography; M, motion.

C. *HFs classification under simulated noise conditions*

When signals were corrupted, the HF sensitivity performance deteriorated (below 75% when using the SC features only), and the multi-sensor approach partially compensated for performance loss (see Fig. 5).

V. Conclusion

The current study shows initial feasibility and advantages for automatically detecting HFs in women, by using a multi-sensor approach and consumer-grade sensors placed on the wrist, a commonly targeted location in the consumer wearable space. We show performance gains, performance differences during wake vs sleep, and robustness to missing sensors. Automatic detection of HFs can provide women and clinicians with new insight into HF patterns, trigger, and impact, advancing the field of women's health.

References

[1] R. R. Freedman, "Physiology of hot flashes," *Am J Hum Biol,* vol. 13, no. 4, pp. 453-64, Jul-Aug, 2001.

[2] R. Bansal, and N. Aggarwal, "Menopausal Hot Flashes: A Concise Review," *J Midlife Health,* vol. 10, no. 1, pp. 6-13, Jan-Mar, 2019.

[3] J. S. Carpenter, M. A. Andrykowski, R. R. Freedman, and R. Munn, "Feasibility and psychometrics of an ambulatory hot flash monitoring device," *Menopause,* vol. 6, no. 3, pp. 209-15, 1999.

[4] M. Forouzanfar, M. Zambotti, A. Goldstone, and F. Baker, "Automatic Detection of Hot Flash Occurrence and Timing from Skin Conductance Activity."

[5] F. Baker, M. Forouzanfar, A. Goldstone, S. Claudatos, H. Javitz, J. Trinder, and M. de Zambotti, "Changes in heart rate and blood pressure across nocturnal hot flashes associated with or without arousal from sleep," *Sleep,* vol. 42, no. 11, pp. pii: zsz175, 2019.

[6] C. Iber, *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*: American Academy of Sleep Medicine, 2007.

[7] E. Štrumbelj, and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowledge and information systems,* vol. 41, no. 3, pp. 647-665, 2014.