A Comparative Study between ECG-based and PPG-based Heart Rate Monitors for Stress Detection

Anuja Pinge, Soumyadip Bandyopadhyay, Surjya Ghosh, Sougata Sen
Department of Computer Science and Information Systems, BITS Pilani Goa Campus.
{p20200052, soumyadipb, surjyag, sougatas}@goa.bits-pilani.ac.in

***Abstract*—Recent advances in the field of wearable sensing has promoted the emergence of many health tracking devices, including heart rate monitors. Heart rate monitors are commonly either chest-based or wrist-based. Currently, it is unclear whether there is a substantial difference in the performance of these different heart rate monitors. To determine the difference in the performance, in this paper, we compare two chest-worn heart rate monitors and one wrist-worn heart rate monitor. Our initial results indicate that there is substantial difference between the devices – the root mean square error between devices can be above 10 beats per minute. However, even though there is difference in performance of different heart rate monitors, yet each of these devices are capable of detecting stress (using an machine learning model) with a F1-score of above 0.8. In this paper, we also introduce the idea of formally verifying the rules obtained from the machine learning classifier; such formal verification will enable improving the explainability and confidence of the outcome of the machine learning models.**

***Index Terms*—Heart Rate Variability, Stress detection, mobile health (mHealth), smartwatches, heart rate monitors**

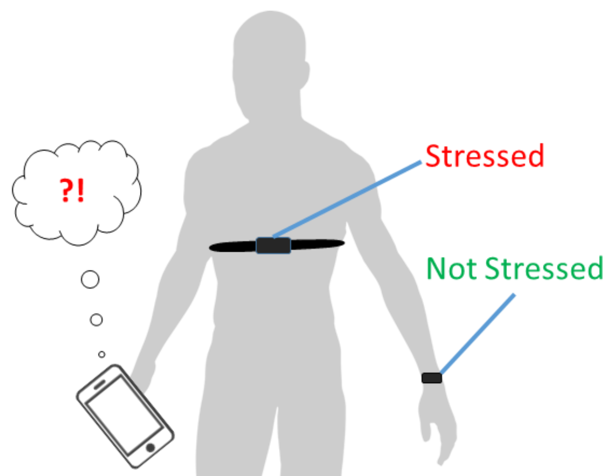

Fig. 1. Different heart rate monitoring devices might output different heart rate values, which in turn might affect the performance of an automatic physiological stress detection algorithm.

I. Introduction

With the advancement in wearable sensing technology, it is gradually becoming possible to monitor various health-related outcomes, one among which is an individual's heart rate. Measuring the heart rate is beneficial for detecting outcomes and issues like arrhythmia [1], the physical activity state of an individual [2], or even their physiological stress levels [3]. Two common approaches for measuring the heart rate and the heart rate variability are via the electrocardiogram (ECG) sensors and the photoplethysmogram (PPG) sensors. Devices such as chest-worn heart rate monitor usually measure continuous heart rate of an individual through ECG, while a wrist-worn device commonly uses a light-based PPG approach to continuously determine one's heart rate measurements.

It is commonly perceived that a chest-worn ECG device provides more accurate heart rate readings as compared to a wrist-worn PPG-based device. In the past, we extracted features from a chest-worn heart rate monitor's data to detect whether an individual was experiencing instantaneous physiological stress [3]. We found that the chest-worn devices performed reasonably well in detecting stress [4]. However, chest-worn heart rate monitoring devices are less commonly used as compared to wrist-worn devices such as smartwatches and fitness bands [5]. Although extracting heart rate data from wrist-worn devices is possible, however, currently researchers have varied opinion about the performance of the wrist-worn devices [6], [7]. Thus, it is currently unknown whether (i) a wrist-worn device can provide similar performance in capturing the raw heart rate data as a chest-worn device, and (ii) can we use the heart rate data extracted from different heart rate monitoring devices for detecting stress under exactly the same context? We pictorially show this in Fig. 1.

This paper aims to determine whether the two category of heart rate monitoring devices – chest-worn heart rate monitors and wrist-worn heart rate monitors – provide same or similar heart rate readings, and whether it is possible to use the heart rate readings provided by each of the devices to determine whether an individual is undergoing physiological stress. To this end, in this paper, we compare three commonly used heart rate monitoring devices – one wrist-worn fitness band, and two chest-worn heart rate monitors. Specifically, we compare the performance of a Garmin Vivosmart 4 Smartwatch,[1] a Polar H10 Heart Rate Monitor,[2] and a Garmin HRM Dual Heart Rate Monitor[3] in two aspects: (i) whether the heart

[1]https://www.garmin.co.in/products/intosports/vivosmart-4-black-large/. Accessed: 22-Nov-2021.
[2]https://www.polar.com/en/products/accessories/H10_heart_rate_sensor. Accessed: 22-Nov-2021.
[3]https://www.garmin.com/en-US/p/649059. Accessed: 22-Nov-2021.

| //$31.00 ©10.1109/COMSNETS53615.2022.9668342

rate readings obtained from these devices are similar, and (ii) whether the three devices' physiological stress detecting performance are similar. We will like the readers to note that in this paper we do not propose any novel stress detection algorithm, or improve any existing stress detection algorithm. Rather, this paper focuses on comparing the difference in heart rate readings and stress prediction (using an machine learning model) for each of the three devices. Additionally, most existing physiological stress detection works rely on empirical evaluation to determine physiological stress. In this paper, we introduce a notion of formalism for the stress detection mechanism. We provide an initial insight about how we can take a formal verification approach to confirm the rules that are empirically determined. Such formal verification of rules will be a stepping stone towards explainability of rules that we empirically obtain for physiological stress detection.

The key contribution of this paper are:

- This paper compares the output of three common, off-the-shelf heart rate monitors. We conducted a user study with 5 participants who wore the three devices while carrying out a laboratory-based study. We compared the heart rate readings obtained from the three devices for the 5 participants and observed that the root mean square error (RMSE) between readings of the two chest-worn heart rate monitors was 5.2 beats per miute (BPM), while the RMSE between the Polar H10 heart rate monitor and the Garmin Vivosmart 4 smartwatch was 10.23 BPM.
- In the laboratory study, participants were introduced to three different types of stressors. We collected sensor data (e.g., heart rate, R-R interval) from these devices and used a machine learning model to detect stress. We observed that the difference in F1-score of detecting stress by the three devices was within 5%. This shows that it is indeed possible to use any of the devices for detecting physiological stress.
- We extracted one rule that was generated by the machine learning model for detecting 'stressed', and one that was used for detecting 'not stressed'. We describe a formal verification approach that can be used to validate the rules generated by the Random Forest based machine learning classifier for determining physiological stress.

II. Background and Related Work

In this section, we first describe how heart rate and stress are commonly measured. Traditionally, heart rate measurement has been ECG-based, where 12 leads are connected to several body parts of an individual. These leads capture the cardiac electrical activity, i.e., the polarization and depolarization of the atrium and the interventricular septum [8]. These electrical activities are commonly represented by the PQRST wave, and the distance between two consecutive waves is measured by measuring the distance between the *R* of the two waves. This distance is called the R-R interval and is measures in milliseconds. The two chest-worn ECG devices that we used in our study output the heart rate data in beats-per-minute, and the R-R interval data in milliseconds. An orthogonal approach for measuring the heart rate and heart rate variability is via the PPG sensor. The PPG sensing mechanism relies on measuring the blood volume change during the cardiac cycle. This measurement is done by projecting a light beam towards an artery – the amount of light absorbed changes based the blood volume. The blood volume in the artery changes based on the cardiac cycle. By measuring the instantaneous blood volume level, one can determine the moment in the cardiac cycle that the PPG sensor is measuring.

Wearable-based heart rate measurement: Capturing the heart rate using the 12 leads ECG approach causes mobility issues. More recently, researchers have worked towards capturing the heart rate data using either wearable devices or infrastructure placed devices [9], [10]. Among wearable devices, researchers have experimented with both ECG sensors [9], as well as PPG sensors [11]. Researchers have explored both custom built devices as well as commercial off-the-shelf devices, and have measured and compared the performance of these devices in various settings [12]. Some have even compared the performance of multiple studies to demonstrate reproducibility [3]. Phan et al. compared the performance of a smartwatch with a pulse oximeter and an ECG device [6]. They observed that the performance of all the devices were similar. Contrary to Phan et al.'s finding, Wang et al. found that the devices have differences in outcomes [7]. None of these studies, however, compared the performance of the devices while inducing stress. In our study, we measure the heart rate data from three devices simultaneously, while inducing physiological stress.

Stress detection: Over the years, several researchers have demonstrated the possibility of detecting stress using numerous approaches [13], including via the heart rate information [14], the speech and galvanic skin response (GSR) [15]. Researchers have performed physiological stress detection studies with various demographic groups. For example, King et al. detected stress levels of pregnant women [16], while Healey and Picard detected the stress level of drivers [12], and Egilmez et al. detected the stress levels of college students [17]. Researchers have explored both in-laboratory data collection, as well as free living data collection [18], [19]. We have currently performed a study similar to Mishra et al.'s study [9]. The stressors in our study (described in Section III) were used in several prior studies. Specifically, the socio-evaluative stressors was used by Hovsepian et al. [19], the mental arithmetic stressor was employed by Mishra at al. [9], while the ice bucket stressor was used by Egilmez et al. [17].

III. Methodology

We next describe the methodology adopted in this study. **Dataset:** For this study, we recruited 7 participants (4 males, 3 females, aged between 25 years and 35 years), either from our lab, or from among the family members of one research team member. Participants were asked to wear the three devices

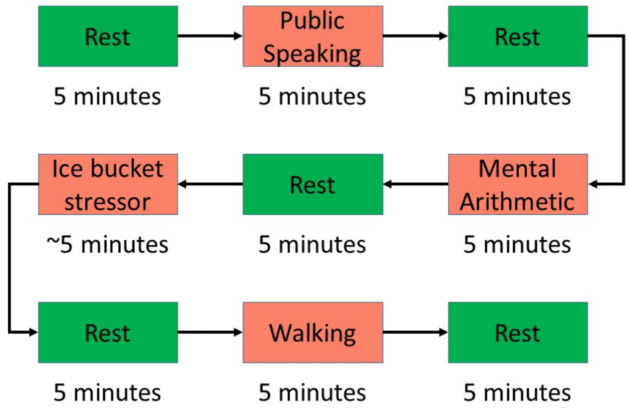

Fig. 2. Series of activities performed by participants in the study protocol

while performing some in-lab activities. The three devices included: (i) Polar H10 Chest Monitor: This is a chest-worn device device capable of collecting heart rate in BPM at a precision of 1 second, and the R-R intervals in milliseconds; (ii) Garmin HRM Dual Chest Monitor: Similar to the Polar H10 Chest Monitor, this device also collects heart rate in BPM; and (iii) Garmin Vivosmart 4 Smartwatch: This is a fitness band that can measure heart rate, blood oxygen level, stress level, sleep and also energy throughout the day.

We discarded data collected from 2 participants (1 male, 1 female) because for one participant, one of the chest-worn heart rate monitor's strap was loose – we did not get any meaningful readings from the participant, while for another participant, the chest-worn device re-positioned itself, resulting into erroneous reading for a part of the study. For the 5 participants, the average duration of the study was 43 minutes(min 41 minutes, max 45 minutes).

Study Protocol and Data Collection Procedure: We collected the data in a laboratory setting. Participants wore the three above-mentioned devices and performed the following tasks: First, we instructed the participants to rest for 5 minutes. As observed by several prior works, including ours, an initial prior of resting removes any residual stress that might remain from a participants' previous tasks. It also helps collecting the baseline heart rate readings. Next, participants were introduced to the following stressors: (i) Socio-evaluative Stressor, where the participant was asked to publicly speak on a topic for 3 minutes. The participant was provided with the topic during the study and we gave them 2 minutes to prepare, (ii) Mental Arithmetic Stressor, where participants were asked to count from 1000 to 1 in steps of -13, and (iii) Physical Stressor, where participants were asked to dip their hand in an ice bucket for up to 5 minutes. The participants were allowed to rest for 5 minutes between each of the stressors. Finally, to collect the heart rate readings during an active phase, we asked the participants to walk briskly for 2 minutes, and fast for 3 minutes. Fig. 2 pictorially depicts the study protocol for data collection process.

We developed an Android-based smartphone application to collect data from each of these wearable devices over a Bluetooth Low Energy (BLE) connection. This application ran on 3 different smartphones and collected the heart rate data and the R-R intervals from each of the devices. We used the LabFront tool[4] to extract the data from the Garmin Vivosmart 4 Fitness Band, while we collected the data from chest monitors i.e. Polar H10 and Garmin HRM Dual over BLE.

Data Cleaning and Processing: We designed our evaluation protocol to utilize readings coming at a per-seconds duration, i.e., one reading per second. The two chest-worn devices provided the heart rate in BPM and the R-R interval between consecutive beats. The Polar H10 generated one reading per second, and thus there was no need of further pre-processing. Unlike the Polar H10 device, the HRM Dual Pro produces two readings per second. We averaged the two readings produced in a second to obtain one single reading every second. Unlike the chest-worn devices, the smartwatch samples the PPG sensor at 50 Hz and produces one heart rate reading every 15 seconds. Additionally, it also provides the R-R intervals. We utilized the R-R interval data to obtain the heart rate reading from the smartwatch, at a granularity of one second.

We first clean the data obtained from these devices. The cleaning process includes removing any invalid data – i.e., readings where the heart rate data is below 55 or above 220 bpm. Additionally, if there are missing readings, we interpolated the data based on previous and subsequent readings. We next pre-processed the cleaned data. To counter the inter-user difference in the heart rate readings, we used the min-max normalization approach to normalize the data. Min-max normalization transforms the heart rate data to a range [0,1]. At the end of the cleaning and pre-processing step, we had 13,054 data instances from the 5 participants.

Feature Extraction and Ground Truth Labeling: We grouped 60 one-second readings from each of the devices into a 1-minute window. There was a 50% overlap between subsequent 1-minute windows. For each 1-minute window, we computed several time-based features from both heart rate data, as well as the R-R interval data. These features were similar to features extracted by Mishra et al. in [9]. Specifically, we computed the maximum heart rate, minimum heart rate, mean heart rate, median heart rate, standard deviation, 80^{th} percentile, 20^{th} percentile features from the heart rate variability data. From the R-R interval data, we computed features similar to the ones for heart rate. However, we computed one additional feature – root mean square of successive differences (RMSSD) for the R-R interval. RMSSD is computed by subtracting subsequent R-R interval values, squaring the value and computing the square root of the mean of the values. The windows extracted from 'Rest' periods were labelled as 'Not Stressed', while readings from the 'Stressor' periods were labelled as 'Stressed'.

[4]https://www.labfront.com/. Accessed 13-Dec-2021.

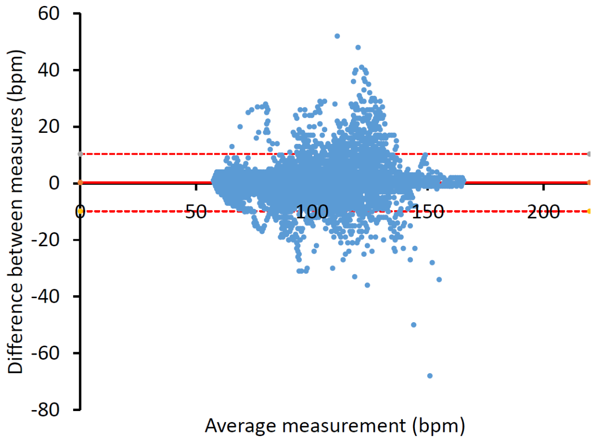

Fig. 3. Comparison of difference of performance of the two chest-worn HRMs in capturing the heart rate reading for all participants. The dotted line indicates the 95% confidence interval of the differences.

Model Construction: We used the above-mentioned features to train a Random Forest model (as implemented in the Scikit-learn library[5]) for stress detection (binary classification). To ensure the generalizability of stress detection, we train a person-independent model. We train separate models for each device for stress detection based on the features extracted (using the data collected from that device only). We obtain output from each of the three devices independently, i.e., each device provides a classification output every 30 seconds (as there is a 50% overlap between windows).

IV. Evaluation

In this section, we discuss the evaluation approach, and we compare the performance in (a) heart rate measurement and (b) stress detection across devices.

Evaluation Approach: Prior work has shown that the output of a Polar heart rate monitor is comparable to heart rate reading obtained from an ECG monitor. Therefore, we considered the Polar H10 as the baseline, and computed the RMSE for the raw heart rate readings obtained from the other two devices.

To compare the performance of the three devices' capability of detecting stress, we performed a leave one person out cross validation (LOPOCV) using the data obtained from each of the devices. Since we do not aim to demonstrate the robustness of our model, but rather compare the performance of the different devices, a LOPOCV performed with data from five participants should suffice. We used precision, recall, f1-score, and accuracy as the evaluation metrics of stress detection for each of the three devices.

Heart Rate Measurement Comparison: We compared the output of heart rate at a per-second granularity for each of the 5 participants. Fig. 3 presents the Bland Altman plot between the Polar H10 and the Garmin HRM Dual. A Bland Altman plot allows visualizing the relationship between data

[5]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. Accessed: 24-Nov-2021.

TABLE I
Root Mean Square Error between the heart rate readings obtained from the Polar H10 device and the other two devices.

Device	RMSE (bpm)
Garmin HRM Dual	5.2
Garmin Vivosmart 4	10.23

TABLE II
Performance of the three devices in detecting stress

Device	Accuracy	Precision	Recall	F1-score
Polar H10	0.85	0.85	0.84	0.85
Garmin HRM Dual	0.81	0.88	0.76	0.82
Garmin Fitness Band	0.83	0.87	0.74	0.80

obtained from two devices [20]. We observe that the two standard deviation limit of agreement between the two devices is within ± 10 bpm. To determine whether they are actually within acceptable limits, we computed the RMSE between the readings of the Polar H10 device and the other two devices. Overall, we observed that the RMSE between the Polar 10 and the Garmin HRM Dual was 5.2 BPM, while the RMSE between the Polar H10 and the Garmin Vivosmart 4 was 10.23 BPM for all the participants. We report this finding in Table I. Although the mean difference of heart rate readings is not high, in future it will be interesting to determine scenarios where the instantaneous difference is high, and if approaches can be taken to reduce the error.

Physiological Stress Detection Comparison: We compare the stress detection performance of the models constructed using the data collected from different devices in Table II. We observe that the F1-score of the three devices are between 0.8 and 0.85. Interestingly, although the smartwatch has a lot of deviation as compared to the Polar H10 device, yet, it is able to detect stress similar to the other devices. This increases the possibility of using smartwatches in future studies for monitoring physiological stress. One must however note that the recall of the smartwatch is low as compared to the other devices. Although the sample size currently might be small to make any conclusive decisions, however, we will have to identify approaches to improve the smartwatch's recall.

V. Formalism

Now that we have a model that has been derived from sensor data to detect physiological stress in an individual, it is important that we can explain why the model takes specific decisions. We next provide an initial intuition about how we can perform a formal verification of the model.

Let v_{R_1} and v_{R_2} be two R values from the PQRST wave of a cardiac cycle. The $R-R$ interval between these two values is denoted by K, i.e., K is of the form $v_{R_1} - v_{R_2}$. For every two R points in a dataset we can get $R-R$ interval. If any of the $R-R$ interval is strictly less than a threshold level, say, BL, which is empirically obtained, then it is possible that the person is asserted to be experiencing stress. Now we characterize stress predicate which takes the

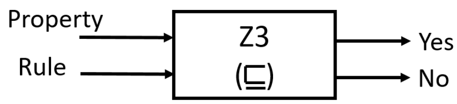

Fig. 4. Framework for formal verification

$R-R$ interval input and gives a Boolean output. Therefore, the stress property is of the form $\forall i, 1 \leq i \leq n, v_{R_i} - v_{R_{i+1}} < BL \rightarrow \forall i, 1 \leq i \leq n \exists v_{R_i} - v_{R_{i+1}} s(v_{R_i} - v_{R_{i+1}})$. Note that we have encoded the stress property using first order predicate calculus semnantics.

Similarly, we can write the 'not-stressed' property, which is the converse of the 'stressed' using the following form $\forall i, 1 \leq i \leq n, v_{R_i} - v_{R_{i+1}} > BL \rightarrow \forall i, 1 \leq i \leq n \exists v_{R_i} - v_{R_{i+1}} \;\; \neg s(v_{R_i} - v_{R_{i+1}})$.

As described in the previous sections, using the data and the machine learning classifier, we get the rules that are generated automatically. Now let us take one rule (one branch from one of the trees in the Random Forest classifier), which is generated automatically. The rule states that if the mean heart rate is greater than 0.35 which is computed by the classifier, and the standard deviation of data in the window is greater 0.05, then the individual is undergoing 'stress'.

From the data one of the stress rule is of the form $avg(HeartRate) > 0.35 \wedge stdev(Heartrate) > 0.05$. Now we characterize $avg(R-R)$. Let val_{R-R} be the function which takes two R picks and generates integer. Formally, we can write as $val_{R-R} \rightarrow \mathbb{R}$. Therefore, we can rewritten $avg(R-R)$ is of the following form $\forall i, 1 \leq i \leq n, \sum_{k=1}^{n} val(v_{R_i} - v_{R_{i+1}})/T$.

Now, using automated theorem prover (Z3),[6] we check the the containment ($\sqsubseteq$) of formal property which is described above and the rules for stress which is generated empirically. If the theorem prover outputs "yes", it implicates the rule satisfies the formal stress property. Similarly, we can check the containment for non-stress property. The overall workflow for formal verification framework is given in Fig. 4. We can use such an approach in the future to determine which rules are effective in determining stress. This will also help explaining the models that we empirically generate.

VI. Discussion and Future work

While we, in this paper, show that it is possible to detect stress from both wrist-worn devices, as well as chest-worn devices with similar performance, however there is still a lot of directions to be explored. We next describe some scope for improvement, and possible future directions.

Generalizability: In our current work we do not focus on developing a generalized stress detection model. Rather, we focus on comparing the difference in performance of different sensing devices in capturing an individual's heart rate and detecting physiological stress. Our current results demonstrates that devices worn on the wrist can also be used to detect stress. Although prior work with other sensors worn on the wrist has demonstrated similar possibilities, however we show that a low-cost commercial off-the-shelf device can perform reasonably too. In future, we aim to collect a larger dataset to ensure the generalizability of our models, as well as confirm the findings that we have presented in this paper. To do that, we will focus on collecting more data not just in laboratory setting, but also in free-living environments.

[6]https://github.com/Z3Prover/z3. Accessed: 24-Nov-2021.

On-device Computation and Interventions: Today's stress detection models collect data from wearable devices or infrastructure devices to detect whether a person is stressed. Most of these stress detection systems work in an offline mode, i.e., they collect data in real time, but process the data offline, usually on a more computationally capable device. However, detecting whether an individual was stressed in future prevents providing necessary interventions during the moment when one is experiencing the physiological stress. Thus, needed are approaches to detect stress in real-time. Real-time detection will allow researchers to introduce just-in-time adaptive interventions (JITAIs) whenever adequate stressful moments are observed.

Recent advances in machine learning, including growing interest in machine learning for tiny embedded devices is opening up the possibility of detecting stress in real-time [21]. Indeed, techniques such as sparsification and quantization allow deploying deep learning models onto the resource constrained devices [22]. Researchers have shown the possibility of deploying deep learning models for sensor streams [23]. In future, we will modify our existing approach to run the stress detection algorithm on resource constrained devices and detect stress in real-time. One challenge with detecting stress in real-time is that motion artifacts affect the stress detection performance. We will augment existing algorithms with IMU sensor data to negate the effect of motion in stress detection.

Energy Requirements: Currently, we utilize all data points collected from heart rate monitors in detecting whether an individual is undergoing stress. Collecting and transmitting continuous data points have severe energy implications. Since these devices are battery powered, continuous collection and transmission of data reduce their battery life. Over the years, researchers have identified techniques such as duty cycling, or piggy-backing to reduce energy. However, currently it is unknown whether duty cycling the heart rate data will affect its performance. In future, we aim to explore approaches to sub-sample the sensor readings obtained from the wearable devices, and determine whether the sub-sampling affects the performance of the stress detection system.

Need for Formal Verification: The outcome from the ML models is often difficult to interpret due to the "black-box" nature of such models. As a result, the widespread usage of these models is challenging in the community. Therefore, to improve the explainability of the ML models, we introduced the formal verification approach based on the rules obtained from the Random Forest classifier. This not only provides explainability of the model, but also provides higher confidence while communicating an outcome to the participants. We, in this paper, have provided an initial formal verification requirement. We envision that the formal verification of the ML-based rules will help improve the models too. We aim to formally verify stress detection models in our future work.

VII. Conclusion

We, in this paper, compared the performance of three heart rate monitors (Polar H10, Garmin HRM Dual , Garmin Vivosmart 4 Fitness Band) that use different technologies (ECG, PPG) for detecting heart rate, and determined their ability to detect physiological stress. We performed a lab-based controlled study involving 7 participants, who encountered a fixed sequence of stressors. We captured the heart rate and R-R interval data from these participants. Our preliminary analysis reveals that the RMSE of heart readings is 5.2, and 10.23 bpm for the HRM Dual Pro, and the Garmin smartwatch respectively with respect to the Polar H10 heart rate monitor. We further trained a person-independent Random Forest-based machine learning model for stress detection from extracted features. The model attained a F1-score of 82%, 79%, and 76% for HRM Dual Pro, Garmin smartwatch, and Polar H10 respectively for stress detection. These findings are further formalized with a symbolic execution engine based on the rules obtained from the ML classifier, to improve the explainability of the model and increase the confidence of the model outcome. Our study shows the possibility of detecting stress using a smartwatch. However, we have to make several improvements before deploying models on smartwatches, which includes providing an explanation about the performance of the model.

Acknowledgements

This paper has benefited from extensive discussions with several people. In particular, we would like to thank Varun Mishra for his valuable inputs, and William Dowie and the team at PhysioQ/LabFront for assisting in extracting the data.

References

[1] R. Fensli, E. Gunnarson, and T. Gundersen, "A wearable ecg-recording system for continuous arrhythmia monitoring in a wireless tele-home-care situation," in *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*. IEEE, 2005, pp. 407–412.

[2] E. E. Dooley, N. M. Golaszewski, and J. B. Bartholomew, "Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices," *JMIR mHealth and uHealth*, vol. 5, no. 3, p. e7043, 2017.

[3] V. Mishra, S. Sen, G. Chen, T. Hao, J. Rogers, C.-H. Chen, and D. Kotz, "Evaluating the reproducibility of physiological stress detection models," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, Dec. 2020. [Online]. Available: https://doi.org/10.1145/3432220

[4] V. Mishra, G. Pope, S. Lord, S. Lewia, B. Lowens, K. Caine, S. Sen, R. Halter, and D. Kotz, "Continuous detection of physiological stress with commodity hardware," *ACM Trans. Comput. Healthcare*, vol. 1, no. 2, apr 2020. [Online]. Available: https://doi.org/10.1145/3361562

[5] F. Sartor, J. Gelissen, R. Van Dinther, D. Roovers, G. B. Papini, and G. Coppola, "Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients," *BMC Sports Science, Medicine and Rehabilitation*, vol. 10, no. 1, pp. 1–10, 2018.

[6] D. Phan, L. Y. Siong, P. N. Pathirana, and A. Seneviratne, "Smartwatch: Performance evaluation for long-term heart rate monitoring," in *International symposium on bioelectronics and bioinformatics (ISBB)*, 2015.

[7] R. Wang, G. Blackburn, M. Desai, D. Phelan, L. Gillinov, P. Houghtaling, and M. Gillinov, "Accuracy of wrist-worn heart rate monitors," *Jama cardiology*, vol. 2, no. 1, pp. 104–106, 2017.

[8] J. W. Hurst, "Naming of the waves in the ecg, with a brief account of their genesis," *Circulation*, vol. 98, no. 18, pp. 1937–1942, 1998.

[9] V. Mishra *et al.*, "The case for a commodity hardware solution for stress detection," in *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers*, 2018, pp. 1717–1728.

[10] I. Shirakami and T. Sato, "Heart rate variability extraction using commodity wi-fi devices via time domain signal processing," in *IEEE EMBS International Conference on Biomedical and Health Informatics*, 2021.

[11] A. Temko, "Accurate heart rate monitoring during physical exercises using ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2016–2024, 2017.

[12] J. Healey and R. Picard, "Smartcar: detecting driver stress," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 4. IEEE, 2000, pp. 218–221.

[13] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *Computing in Cardiology*, 2011.

[14] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: a meta-analysis and review of the literature," *Psychiatry investigation*, vol. 15, no. 3, p. 235, 2018.

[15] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 209–214.

[16] Z. D. King, J. Moskowitz, B. Egilmez, S. Zhang, L. Zhang, M. Bass, J. Rogers, R. Ghaffari, L. Wakschlag, and N. Alshurafa, "Micro-stress ema: A passive sensing framework for detecting in-the-wild stress in pregnant mothers," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 3, no. 3, pp. 1–22, 2019.

[17] B. Egilmez, E. Poyraz, W. Zhou, G. Memik, P. Dinda, and N. Alshurafa, "Ustress: Understanding college student subjective stress using wrist-based passive sensing," in *IEEE International Conference on Pervasive Computing and Communications Workshops*, 2017.

[18] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: in laboratory and real life," in *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*, 2016, pp. 1185–1193.

[19] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cstress: towards a gold standard for continuous stress assessment in the mobile environment," in *ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 493–504.

[20] D. Giavarina, "Understanding bland altman analysis," *Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.

[21] R. David *et al.*, "Tensorflow lite micro: Embedded machine learning on tinyml systems," *arXiv preprint arXiv:2010.08678*, 2020.

[22] S. Bhattacharya and N. D. Lane, "Sparsification and separation of deep learning layers for constrained resource inference on wearables," in *ACM Conference on Embedded Network Sensor Systems*, 2016, pp. 176–189.

[23] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, and F. Kawsar, "Multimodal deep learning for activity and context recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–27, 2018.