

Stress Detection in Wearables

Alice Liao, Anna Darwish, Bernardo Martinez

9/30/2020

1 Introduction

In the past few years, the use of wearables has exploded [1]. With it, researchers around the globe are exploring how tracking health information via wearables, such as a person’s heart rate, can help gain insight into their physical and emotional states [2]. Across many different functionalities, such as tracking workouts or monitoring sleep, wearables also present an exciting opportunity to detect stress. In doing so, wearable users today, such as Apple watch user Jason Hiner, have become more cognizant of their emotional state and worked to improve it [3]. Given that stress can lead to long term health issues, such as heart disease, high blood pressure, diabetes, and depression or anxiety [4], it is critical for us to ensure that wearables can accurately detect stress and distinguish it from other affective states.

However, researchers have found evidence to be skeptical of the accuracy and insights a wearable watch is capable of offering [5]. For example, experts at the Cleveland Clinic found that heart-rate data in many popular wearables can be wrong 10-20% of the time [6]. While research has been done to examine whether these sensors are capable of detecting excited affective states, such as stress, relative to a subject’s baseline state, there are far fewer studies that examine whether these sensors are capable of distinguishing excited affective states from one another, such as stress vs amusement [7]. A recent paper by Schmidt et. al was able to generate highly accurate machine learning models that classified wearers’ states among baseline, stress, and amusement conditions using an *Empatica E4* wristband [7]. However, these methods were unable to provide significant insight into the relationship between the most informative metrics and these affective states.

In this paper we hope to expand upon the work done within the Schmidt et. al paper by delving deeper into the relationship between the metrics collected by these sensors and the wearers’ affective states with a logistic regression model. Specifically, we hope to accomplish the following goals:

1. Determine whether sensor data are useful in discriminating between stress and amusement conditions
2. Understand and describe how sensor data are useful in discriminating between stress and amusement
3. Determine which types and combination(s) of sensor data are most useful in discriminating between stress and amusement
4. Determine whether we can detect stress (vs amusement) only using the wrist-worn wearable
5. Quantify the heterogeneity across individuals in the response to stress versus amusement

2 Data

2.1 Collection Overview

For the purposes of this paper, we will use the biological definition of stress: an affective state defined to be the ‘nonspecific response of the body to any demand upon it’ [7]. The data used in this paper were collected by attaching a *RespiBAN Professional* chest device and *Empatica E4* wristband to 17 graduate students. Due to sensor malfunctions, two of the subjects’ data had to be discarded, for a final total of 15 subjects. Three of these subjects were women while the rest were men, and their ages ranged from 25 years old to 35 years old. This experiment sought to elicit three different affective states: neutral, stress, and amusement, and accomplished this by having the subjects perform stressful tasks (e.g public speaking) and amusing tasks (e.g watching funny videos) with breaks of meditation between sessions to restore subjects to their baseline. The stress-condition tasks and amusement-condition tasks were interchanged to avoid effects of order [7].

In the Appendix, you can see a full data dictionary in Table 5 that defines the metrics, their sampling rates, and any extracted features. Finally, it is important to note that for Subject 3, there appears to have been a brief malfunction (over a span of approximately 0.1 seconds during stress condition) that led to impossible measurements, such as temperatures of -273°C . We chose to treat these ranges as missing values and used linear interpolation to replace these faulty measurements.

2.2 Exploratory Data Analysis

We plotted the distributions of predictors for each of the subjects and realized for some people, certain metrics serve as perfect separators for amused vs. stressed conditions (i.e. the distributions under the two conditions are significantly different or separated), but that's not always the case. Among all predictors, electrodermal activity (EDA) and temperature are the two physiological signals that appear to be separators for most of the subjects.

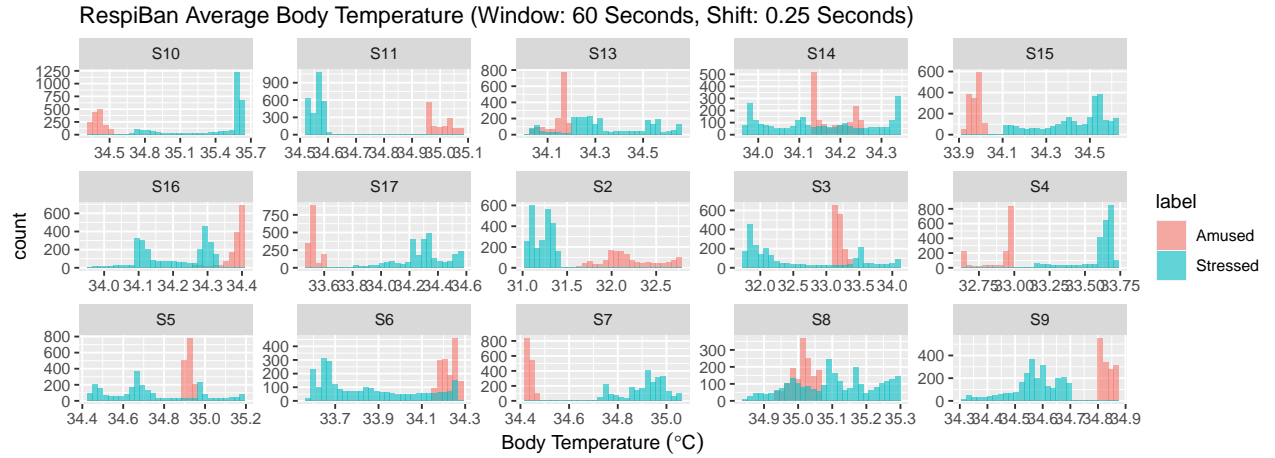


Figure 1

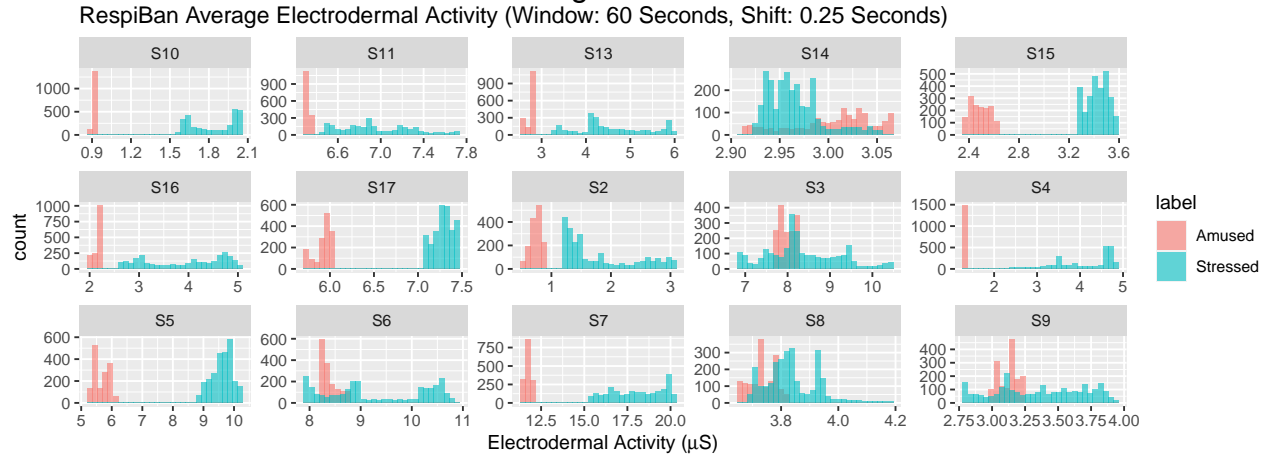


Figure 2

From Figure 1 we see body temperatures are different in amused vs stressed states. For some subjects, such as Subject 10 and 11, the difference is more pronounced in that the distributions of body temperatures are completely separated by the two states. Similarly in Figure 2 we observe different electrodermal activities in different emotional states. The histograms also suggest not everyone reacts to amusement and stress in the same way. In general, it seems higher EDA values were associated with stress, but for Subject 14, some of the higher EDA values occurred during the amusement conditions. As for temperature, some subjects had higher temperature under stress (e.g. Subject 10 and Subject 15) whereas some had lower temperature

when amused (e.g. Subject 11 and Subject 17). Skin temperature and EDA from the *Empatica E4* appear to be separators too, and their distributions can be found in Appendix.

3 Methods

Our first step with this dataset involved aligning the *RespiBAN* data and *Empatica E4* data. For the *RespiBAN* data, each metric was collected at 700Hz. Across all metrics, 700Hz was well above the minimum sampling rate [8], [9], [10], [11], which is defined to be the minimum number of samples needed to glean precise feature extractions. For more detail on the minimum sampling rates (MSRs) across the different metrics, please reference Table 5 in the Appendix. Despite these different MSRs, these metrics ultimately needed to be aligned to represent distinct and whole samples. We selected to downsample to 4 Hz, which has been documented as the minimum sampling rate of electrodermal activity (EDA). From our data exploration, EDA appeared to be a key metric in distinguishing stress and amusement, so we chose to prioritize it when downsampling.

We accomplished this downsampling by calculating our features over a window of one minute with a shift of 0.25 seconds. This was motivated by the Schmidt et. al’s paper’s window and shift decisions. We conducted sensitivity analysis by recreating our final chest and wrist models using a 1 minute window with a 0.5 second shift, along with a 2 minute window with a 30 second shift. We found our results to be generally consistent for both. Please reference Tables 6-9 for these results.

For this paper, we have decided to generate several logistic regression models to better understand and interpret the relationship between these metrics. We began by exploring histograms and scatterplots of our metrics. From this we started with a model with the following predictors: RespiBAN EDA, RespiBAN Body Temperature, and Subject. Including Subject ensures independence among our samples, as the observations within the same subject were highly correlated.

When we were exploring models for the wrist data, we initially began with the mean EDA as a predictor, as it perfectly separated the stress and amused condition for many of the subjects (see Appendix). However, this caused significant model instability and severely inflated the standard errors. For that reason, we explored other statistical features of EDA and noticed that the standard deviation of EDA also performed quite well. For subjects such as Subjects 2 and 15, who did not have as strong of separation when only using the standard deviation of EDA, we noticed the mean wrist temperature served as a useful metric. Finally, for quantifying heterogeneity, we explored interaction effects with skin temperature.

To validate our models, we used confusion matrices, ROC curves, and binned deviance residual plots to evaluate our model assumptions. Finally, we ran 10-fold cross-validation on each predictive model to estimate the test error.

4 Results

4.1 Chest & Wrist Model

Table 1: Chest Model Output

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	99.22	1.914	51.84	0
EDA_CHEST_MEAN	7.571	0.0744	101.8	0
Temp_CHEST_MEAN	-3.38	0.06187	-54.63	0
subjectS10	8.976	0.2203	40.73	0
subjectS11	-30.75	0.3427	-89.75	0
subjectS13	-8.247	0.1951	-42.27	0
subjectS14	-5.671	0.1698	-33.39	2.07e-244
subjectS15	-5.734	0.18	-31.86	8.25e-223
subjectS16	-2.688	0.1945	-13.82	1.956e-43
subjectS17	-34.27	0.3745	-91.52	0
subjectS3	-47.56	0.4782	-99.46	0
subjectS4	-3.53	0.2635	-13.4	6.254e-41
subjectS5	-38.14	10.32	-3.696	0.0002187
subjectS6	-47.5	0.48	-98.95	0
subjectS7	-86.71	23.48	-3.693	0.0002218
subjectS8	-8.85	0.2111	-41.91	0
subjectS9	-5.575	0.188	-29.66	2.863e-193

$$\log\left(\frac{P_i(Stress)}{1 - P_i(Stress)}\right) = \beta_0 + \beta_1 \overline{EDA}_i^{RB} + \beta_2 \overline{TEMP}_i^{RB} + \sum_{j \in S} \beta_j I(Subject = Subject_j)$$

$$S = \{3, 4, \dots, 11, 13, \dots, 17\}$$

Note that the *RB* superscript indicates that these measures came from the RespiBAN, and the baseline subject for comparison was Subject 2.

For every 0.1 μs increase in EDA from the RespiBAN, we expect that the odds that the subject is stressed multiplies by a factor of 2.132. For every 0.01 degree Celsius increase in body temperature, we predict that the likelihood the subject is stressed decreases by a factor of 0.967. We will not be interpreting the subject coefficients as they were included in the model for independence purposes.

This model also provided statistically significant values for all the coefficients used.

Chest	Amused	Stressed
Amused	18285	2547
Stressed	4015	37317

From the confusion matrix, we can see that we are slightly more likely to produce a false positive with this model, i.e. classifying someone as stressed when they are amused. This is likely due to the fact that we have about 1.8 times as many stress-condition samples than amuse-condition samples.

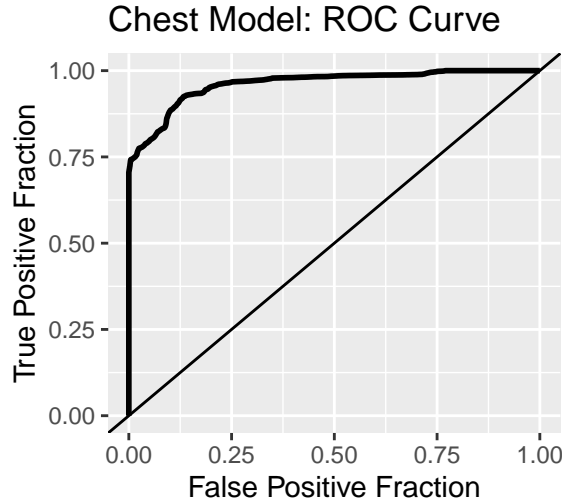
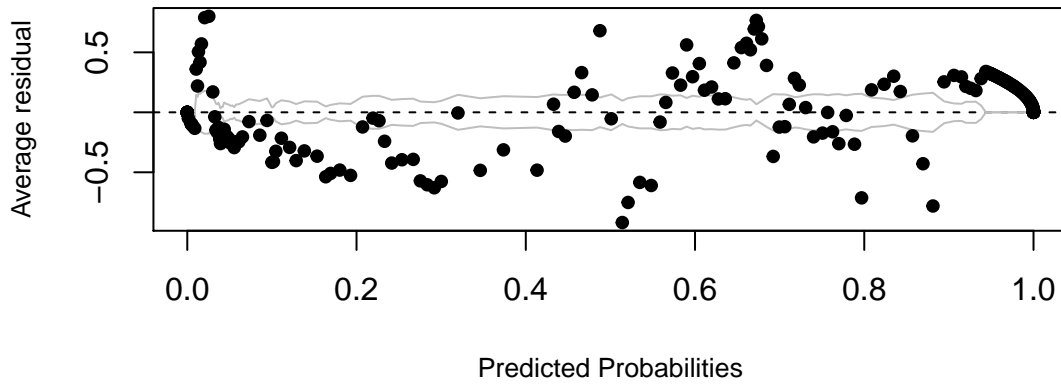


Figure 3

Figure 4: Binned Residual vs. Predicted Probabilities



From the ROC curve in Figure 3, we achieved an AUC of approximately 0.96, which indicates that our model has a relatively high accuracy for predicting the response.

For the binned residuals vs predicted probabilities in Figure 4, we see that there appears to be a strong cubic curve. While this may suggest that we should take our predictors to a higher power, we could not find any literature to motivate such a decision. Instead, we believe there are external factors that these metrics could not capture in this model. Please reference the Appendix for the remaining binned residual plots. When examining the binned residuals against the mean EDA, we see that beyond $9 \mu S$, the residuals go to 0. This is likely due to the perfect separation we saw in our data exploration. Before this, the scatter appears to be random about the 0 line. Finally, the binned residuals against body temperature show that we are generally incurring false negatives on the extreme ends of body temperature and incurring false positives between $33^{\circ}C$ and $35^{\circ}C$.

After performing 10-fold cross validation, the average accuracy rate on the test sets was 0.894 with the worst accuracy rate being 0.89 and the best accuracy rate being 0.898. In contrast, if we had naively predicted the most frequent class in the test set, our mean accuracy rate was 0.641 with the worst accuracy rate being 0.635 and the best accuracy rate being 0.648.

	Estimate	Std. Error	z value	Pr(> z)
--	----------	------------	---------	----------

Table 3: Wrist Model Output

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.68	0.6061	-4.421	9.819e-06
EDA_EMP4_STDDEV	31.07	0.3254	95.47	0
TEMP_EMP4_MEAN	0.06455	0.01794	3.598	0.0003202
subjectS10	-0.2773	0.05088	-5.45	5.033e-08
subjectS11	-1.208	0.05826	-20.74	1.556e-95
subjectS13	-7.223	0.09453	-76.41	0
subjectS14	1.028	0.05248	19.59	1.971e-85
subjectS15	0.3081	0.07922	3.889	0.0001006
subjectS16	0.1272	0.07657	1.661	0.09663
subjectS17	0.7162	0.05213	13.74	5.979e-43
subjectS3	-1.297	0.07351	-17.65	1.033e-69
subjectS4	-1.025	0.06266	-16.35	4.256e-60
subjectS5	-1.384	0.05362	-25.82	5.527e-147
subjectS6	-3.485	0.07044	-49.48	0
subjectS7	-1.31	0.05567	-23.53	1.899e-122
subjectS8	0.3233	0.05079	6.367	1.932e-10
subjectS9	0.6282	0.05437	11.55	7.005e-31

$$\log\left(\frac{P_i(Stress)}{1 - P_i(Stress)}\right) = \beta_0 + \beta_1 * SD(EDA_i)^{E4} + \beta_2 * \overline{TEMP}_i^{E4} + \sum_{j \in S} \beta_j * I(Subject = Subject_j)$$

$$S = \{3, 4, \dots, 11, 13, \dots, 17\}$$

Note that the $E4$ superscript indicates that these measures came from the *Empatica E4* wristband, and we have kept the same baseline with Subject 2.

For every $0.1 \mu s$ increase in standard deviation of the EDA from the *Empatica E4*, we expect that the odds that the subject is stressed multiplies by a factor of 22.348. For every 0.01 degree Celsius increase in skin temperature, we predict that the likelihood the subject is stressed decreases by a factor of 1.001. Again, we will not be interpreting the subject coefficients per the aforementioned reason.

Wrist	Amused	Stressed
Amused	14554	1481
Stressed	7746	38383

From the confusion matrix, we are again slightly more likely to produce a false positive with this model.

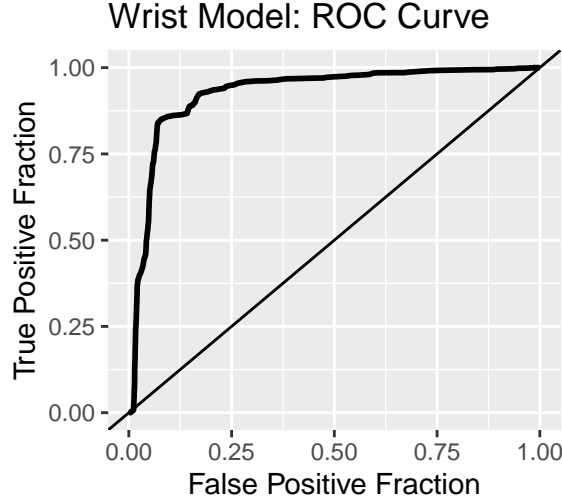
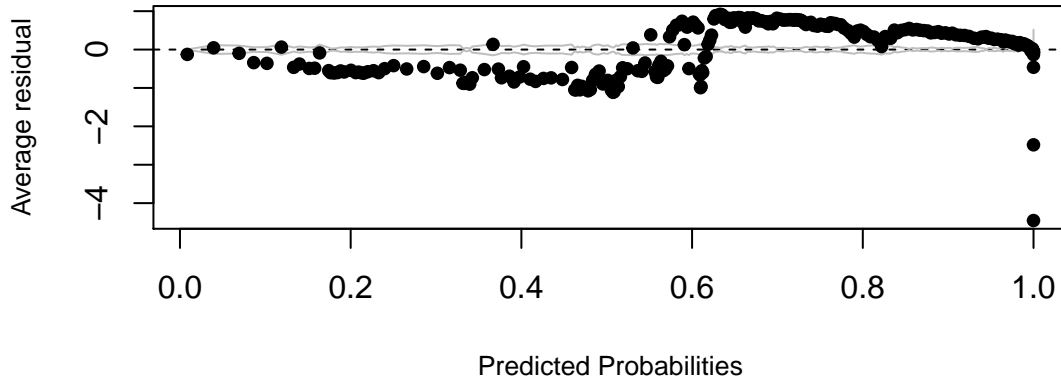


Figure 5

Figure 6: Binned Residual vs. Predicted Probabilities



From the ROC curve in Figure 5, we achieved an AUC of approximately 0.9273, which indicates that our model has a relatively high accuracy for predicting the response.

For the binned residuals vs predicted probabilities, we see that there appears to be another strong cubic curve. Again, while this may suggest that we should take our predictors to a higher power, we could not find any literature to motivate such a decision among the predictors in this model. Please reference the Appendix for the remaining binned residual plots. When examining the binned residuals against the standard deviation of the EDA, we see another strong curve along with significantly low residuals. The reason for this may be due to convergence issues that resulted from perfect separation for some of the subjects. Finally, the binned residuals against skin temperature show a generally random scatter, excluding a brief interval between 31°C and 32°C. Additionally, we also have a few significantly large residuals towards 35°C.

After performing 10-fold cross validation, the average accuracy rate on the test sets was 0.852 with the worst accuracy rate being 0.843 and the best accuracy rate being 0.858. In contrast, if we had naively predicted the most frequent class in the test set, our mean accuracy rate was 0.54 with the worst accuracy rate being 0.53 and the best accuracy rate being 0.555.

4.2 Subject Heterogeneity

We noticed that some of the subjects’ likelihood of being stressed had opposite relationships with respect to their recorded skin temperature from the wristband. For example, Subject 6 recorded higher skin temperatures during his stress condition while Subject 4 recorded lower skin temperatures during his stress condition. Please reference Table 10 in the Appendix for this model, which includes mean skin temperature and an interaction term with subject. In this summary output of our interaction model, you will note that Subject 6’s positive interaction effect with temperature counteracts the main effect of temperature on the likelihood of stress. Specifically, for a 0.01 increase in degrees Celsius, we expect that the likelihood that Subject 6 is stressed multiplies by a factor of 1.604. However for Subject 4, we expect that for a 0.01 increase in degrees Celsius, their likelihood of being stressed decreases by a factor of 0.525.

5 Discussion

Our first goal was to determine whether sensor data are useful in discriminating between stress and amusement conditions. Our estimated accuracy from our first model, which was approximately 0.894, was significantly above the accuracy resulting from naively selecting the most frequent class. For this reason, we have strong evidence to believe sensor data are useful in distinguishing these conditions.

In order to address our second and third goal, which was to determine how and which sensor data are useful in this task, we studied physiological literature discussing stress, conducted feature extraction, and created a logistic regression model motivated by our data analysis. From this, we saw that the mean electrodermal activity and body temperature were significant predictors and interpreted the resulting coefficients.

Next, we examined whether the wristband sensor data from the *Empatica E4* was also useful in discriminating between the conditions alone, as people are unlikely to consistently wear heavy duty chest sensors. While we faced the limitation of inflated standard errors and convergence challenges that prevented us from using a seemingly useful predictor, we were still able to produce a model with an average accuracy of 0.852. Again, this was significantly above the accuracy resulting from naively selecting the most frequent class. While this accuracy is unsurprisingly slightly lower than that of the chest model, it is clearly still a useful indicator.

Lastly, an important question to answer is the heterogeneity across individuals in the response to stress versus amusement, as physiological responses to stress can differ significantly from person to person. We noticed in our exploratory data analysis that there was an opposite relationship in wrist temperature for some of the subjects. We quantified this difference by creating another logistic regression model and comparing the expected change in the odds as skin temperature increased for Subject 4 and Subject 6. From this, we saw that higher skin temperatures were associated with lower odds of being stressed for Subject 4 and higher odds for Subject 6.

6 Limitations and Looking Ahead

Several of the limitations in the analysis include our inability to cater to each metric’s preferred sampling rate, the limited number of affective states, the imbalance of conditions in the dataset, the imprecision in the devices themselves, the lack of representation from younger and older age groups, and the limited representation of women. One possible improvement of this dataset could have been multiple stress trials or conditions. For future studies, it will be necessary to examine whether sensor data from wearables is capable of distinguishing other affective states that yield similar physiological responses, such as anger, and to include subjects from a broad range of ethnicities, races, and age groups. Another possible avenue to explore would be testing popular consumer wrist sensors that claim to accurately detect emotion. Finally, a possible improvement to this study could have been conducting nested cross-validation, in which the hyperparameters would be window size/shift, sampling frequency, etc, so a wider range of views on the data are explored.

7 Appendix

7.1 Tables

Table 5: Data Dictionary

Metric	Description	MSR	Features
RespiBAN Chest Sensor			
ECG (mV)	Electrocardiography: 700 Hz	50 Hz	Mean Std Deviation
EDA (μs)	Electrodermal Activity: 700 Hz	4 Hz	Mean Std Deviation Max
EMG (mV)	Electromyography: 700 Hz	30 Hz	Mean Std Deviation Median
Temp (C°)	Body Temperature: 700 Hz	1 Hz	Mean Std Deviation Median
Resp (%)	Respiration: 700 Hz	10 Hz	Mean Std Deviation
Empatica E4 Wrist Sensor			
BVP	Blood Volume Pulse: 64 Hz	1 Hz (HR)	Mean Std Deviation HR Mean HR Std Deviation
EDA (μs)	Electrodermal Activity: 4 Hz	4 Hz	Mean Std Deviation Max
Temp (C°)	Skin Temperature: 4 Hz	1 Hz	Mean Std Deviation Median

Note that while both sensors also included accelerometer data, it was not included in this analysis as it was only used to align the data points and distinguish the timestamps of the amusement and stress states. Other studies also have considered this variable and ultimately found it not to be useful, as you can read here [9].

For our final model, we decided to use linear interpolation to upsample the the electrocardiography *Empatica E4* data [8]. For example, say we observe the values [2.0, 6.0, 12.0, 24.0] with 2 Hz as the current sampling rate (i.e. 2 and 6 are observations from the first second), and we wish to upsample to 4 Hz. We would translate these values to [2.0, 3.33, 4.66, 6.0, 12.0, 16.0, 20.0, 24.0]. We tested our model's sensitivity to this decision by also using forward projection (in the example above, forward projection would yield [2.0, 2.0, 6.0, 6.0, 12.0, 12.0, 16.0, 16.0]) and saw similar results, which is located in Table ** in the Appendix (REMOVE IF FALSE).

Table 6: Chest Model Output: 1 Minute Window, 0.5s Shift

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	99.23	2.708	36.65	4.398e-294

	Estimate	Std. Error	z value	Pr(> z)
EDA_CHEST_MEAN	7.571	0.1052	71.96	0
Temp_CHEST_MEAN	-3.38	0.08752	-38.62	0
subjectS10	8.976	0.3117	28.8	2.243e-182
subjectS11	-30.76	0.4847	-63.46	0
subjectS13	-8.245	0.2759	-29.88	3.665e-196
subjectS14	-5.671	0.2403	-23.6	3.691e-123
subjectS15	-5.734	0.2546	-22.53	2.247e-112
subjectS16	-2.686	0.2751	-9.766	1.583e-22
subjectS17	-34.27	0.5296	-64.72	0
subjectS3	-47.56	0.6763	-70.33	0
subjectS4	-3.523	0.3717	-9.479	2.567e-21
subjectS5	-38.14	14.58	-2.616	0.008897
subjectS6	-47.5	0.6788	-69.97	0
subjectS7	-86.71	33.28	-2.605	0.009175
subjectS8	-8.849	0.2987	-29.63	6.029e-193
subjectS9	-5.575	0.2659	-20.97	1.363e-97

Table 7: Chest Model Output: 2 Minute Window, 30s Shift

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	126.2	24.06	5.248	1.536e-07
EDA_CHEST_MEAN	8.741	0.9775	8.942	3.836e-19
Temp_CHEST_MEAN	-4.275	0.7811	-5.473	4.434e-08
subjectS10	11.66	2.852	4.089	4.333e-05
subjectS11	-34.23	4.34	-7.888	3.066e-15
subjectS13	-8.41	2.496	-3.369	0.000755
subjectS14	-5.687	2.15	-2.645	0.008163
subjectS15	-5.677	2.283	-2.486	0.0129
subjectS16	-2.05	2.586	-0.7927	0.428
subjectS17	-38.7	4.842	-7.992	1.328e-15
subjectS3	-54.71	6.268	-8.728	2.6e-18
subjectS4	-3.378	3.931	-0.8593	0.3902
subjectS5	-43	359.2	-0.1197	0.9047
subjectS6	-53.65	6.179	-8.683	3.868e-18
subjectS7	-99.43	694.3	-0.1432	0.8861
subjectS8	-8.924	2.612	-3.417	0.0006332
subjectS9	-5.211	2.348	-2.219	0.0265

For the chest model with a 1 minute window and 0.5 second shift, the estimates for the mean EDA and mean body temperature coefficients from the *RespiBAN* are nearly identical to our final model. Further, for the chest model with a 2 minute window and 30 second shift, the estimates for the same coefficients were within 2 standard errors of our final model.

Table 8: Wrist Model Output: 1 Minute Window, 0.5s Shift

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.695	0.8571	-3.144	0.001667
EDA_EMP4_STDDEV	31.06	0.4601	67.51	0
TEMP_EMP4_MEAN	0.06501	0.02537	2.563	0.01039
subjectS10	-0.2776	0.07195	-3.858	0.0001145

	Estimate	Std. Error	z value	Pr(> z)
subjectS11	-1.208	0.08239	-14.66	1.204e-48
subjectS13	-7.221	0.1337	-54.02	0
subjectS14	1.028	0.07422	13.85	1.22e-43
subjectS15	0.3098	0.112	2.765	0.005688
subjectS16	0.1285	0.1083	1.187	0.2353
subjectS17	0.7164	0.07372	9.718	2.531e-22
subjectS3	-1.296	0.104	-12.47	1.065e-35
subjectS4	-1.024	0.08861	-11.56	6.587e-31
subjectS5	-1.385	0.07582	-18.26	1.701e-74
subjectS6	-3.484	0.09959	-34.98	4.911e-268
subjectS7	-1.309	0.07873	-16.63	3.998e-62
subjectS8	0.3233	0.07182	4.502	6.736e-06
subjectS9	0.6284	0.07688	8.173	3.007e-16

Table 9: Wrist Model Output: 2 Minute Window, 30s Shift

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.6959	6.446	0.108	0.914
EDA_EMP4_STDDEV	17.13	2.134	8.03	9.782e-16
TEMP_EMP4_MEAN	-0.02851	0.1909	-0.1493	0.8813
subjectS10	-0.3695	0.5525	-0.6689	0.5036
subjectS11	-1.116	0.603	-1.85	0.06424
subjectS13	-6.781	1.02	-6.65	2.925e-11
subjectS14	0.6421	0.5627	1.141	0.2538
subjectS15	-0.1352	0.8452	-0.16	0.8729
subjectS16	-0.2952	0.8225	-0.3589	0.7197
subjectS17	0.4877	0.5644	0.8641	0.3876
subjectS3	-1.702	0.7918	-2.15	0.03159
subjectS4	-0.8217	0.6297	-1.305	0.1919
subjectS5	-1.211	0.5736	-2.111	0.03476
subjectS6	-2.53	0.6755	-3.745	0.0001801
subjectS7	-1.131	0.5748	-1.968	0.04909
subjectS8	0.1861	0.549	0.3389	0.7347
subjectS9	0.3781	0.5821	0.6495	0.516

For the wrist model with a 1 minute window and 0.5 second shift, the estimates for the mean EDA and mean body temperature coefficients from the *Empatica E4* are nearly identical to our final model. However, for the rist model with a 2 minute window and 30 second shift, the estimates for the mean EDA were not within 2 standard errors, and this model did regard mean skin temperature as significant. This may be due to the window size and shift being too large to capture critical changes.

Table 10: Wrist Model Output

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1989	138.3	14.38	7.227e-47
TEMP_EMP4_MEAN	-58.4	4.061	-14.38	6.78e-47
subjectS10	-769.1	152.9	-5.029	4.931e-07
subjectS11	-3980	255429	-0.01558	0.9876
subjectS13	1912	300294	0.006369	0.9949

	Estimate	Std. Error	z value	Pr(> z)
subjectS14	-4336	377504	-0.01149	0.9908
subjectS15	148	168.4	0.8787	0.3796
subjectS16	-4709	405952	-0.0116	0.9907
subjectS17	302730	2096143	0.1444	0.8852
subjectS3	-3599	263851	-0.01364	0.9891
subjectS4	8543	971.2	8.797	1.411e-18
subjectS5	4478	563351	0.007949	0.9937
subjectS6	-3522	161.7	-21.78	3.624e-105
subjectS7	-1349	140.8	-9.579	9.792e-22
subjectS8	355.1	339418	0.001046	0.9992
subjectS9	-2955	185840	-0.0159	0.9873
TEMP_EMP4_MEAN:subjectS10	22.56	4.49	5.024	5.049e-07
TEMP_EMP4_MEAN:subjectS11	120.3	7921	0.01519	0.9879
TEMP_EMP4_MEAN:subjectS13	-53.99	8655	-0.006238	0.995
TEMP_EMP4_MEAN:subjectS14	131	11646	0.01125	0.991
TEMP_EMP4_MEAN:subjectS15	-12.31	5.158	-2.387	0.017
TEMP_EMP4_MEAN:subjectS16	147.8	13323	0.0111	0.9911
TEMP_EMP4_MEAN:subjectS17	-9261	64106	-0.1445	0.8851
TEMP_EMP4_MEAN:subjectS3	109.6	8362	0.01311	0.9895
TEMP_EMP4_MEAN:subjectS4	-263.4	29.65	-8.884	6.472e-19
TEMP_EMP4_MEAN:subjectS5	-130	16419	-0.007919	0.9937
TEMP_EMP4_MEAN:subjectS6	105.7	4.813	21.96	7.65e-107
TEMP_EMP4_MEAN:subjectS7	39.07	4.137	9.443	3.636e-21
TEMP_EMP4_MEAN:subjectS8	-12.48	10289	-0.001213	0.999
TEMP_EMP4_MEAN:subjectS9	88.56	5762	0.01537	0.9877

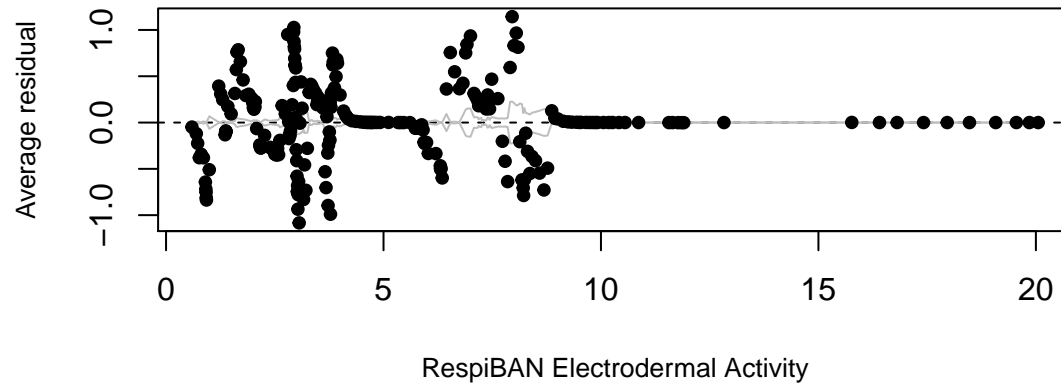
7.2 Plots

7.2.1 Distributions of EDA and Temperature Measured at Wrist

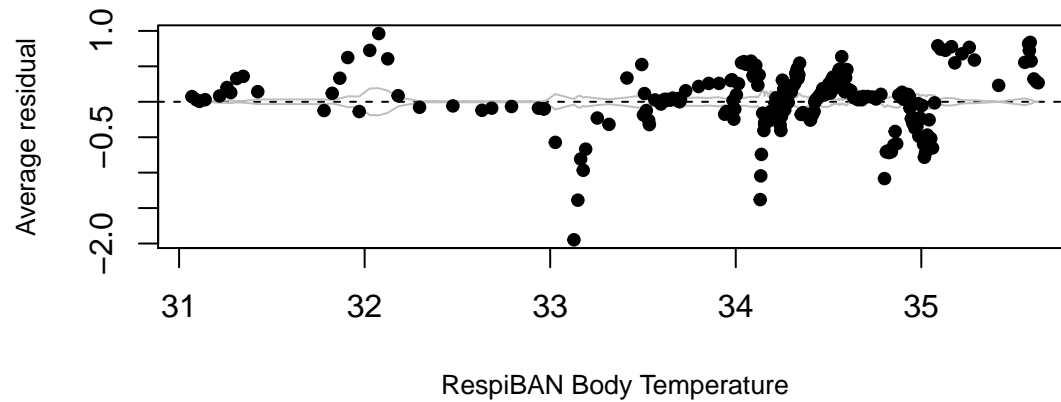


7.2.2 Chest Binned Residual Plots

Binned Residual vs. RespiBAN Electrodermal Activity

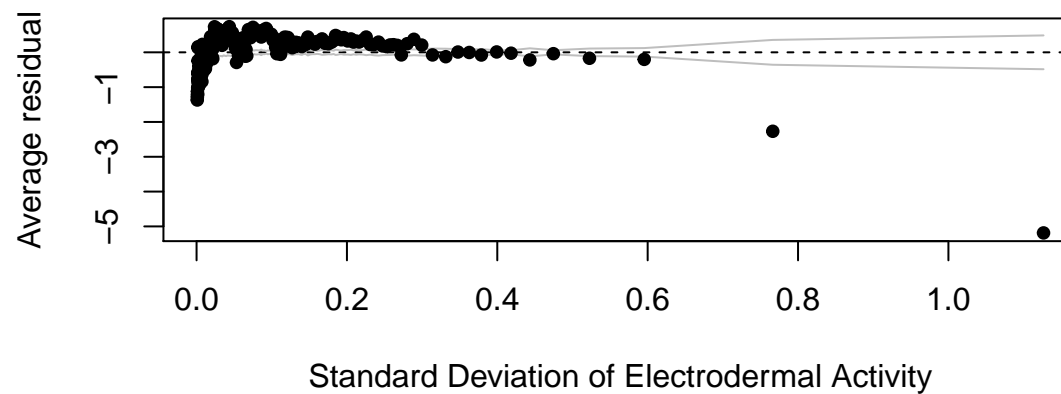


Binned Residual vs. RespiBANBody Temperature

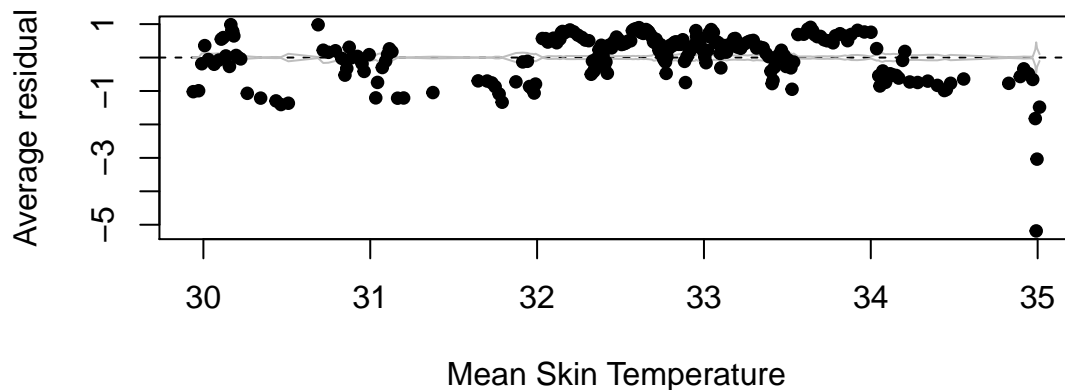


7.2.3 Wrist Binned Residual Plots

Binned Residual vs. EMP4 Electrodermal Activity



Binned Residual vs. EMP4 Skin Temperature



7.3 Bibliography

1. Wray, S. (2017, July 27). Explosion of smart, wearable devices will triple revenue by 2016. Retrieved September 20, 2020, from <https://inform.tmforum.org/features-and-analysis/2014/09/explosion-smart-wearable-devices-will-triple-revenue-2016/>
2. Lu, T., Fu, C., Ma, M., Fang, C., & Turner, A. (2016, September 14). Healthcare Applications of Smart Watches. A Systematic Review. Retrieved September 20, 2020, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052554/>
3. Hiner, J. (2020, April 27). How Apple Watch helped fix my toxic level of stress. Retrieved September 20, 2020, from <https://www.cnet.com/news/apple-watch-uncovered-my-toxic-stress-level-and-helped-me-fix-it/>
4. 5 Things You Should Know About Stress. (n.d.). Retrieved September 20, 2020, from <https://www.nimh.nih.gov/health/publications/stress/index.shtml>
5. Clark, E. (2019, October 17). New Report Finds That More Than Half of Wearable Users Fear Inaccurate Health Data, Malfunctions; Users Cautioned Against Relying on Devices. Retrieved September 20, 2020, from <https://www.prnewswire.com/news-releases/new-report-finds-that-more-than-half-of-wearable-users-fear-inaccurate-health-data-malfunctions-users-cautioned-against-relying-on-devices-301123488.html>
6. Weintraub, K. (2016, October 13). Wearable health monitors not always reliable, study shows. Retrieved September 20, 2020, from <https://www.usatoday.com/story/news/2016/10/12/wearable-health-monitors-not-always-reliable-study-shows/91922858/>
7. Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). Association for Computing Machinery, New York, NY, USA, 400-408. DOI:<https://doi.org/10.1145/3242969.3242985>
8. <https://www.cs.toronto.edu/~guerzhoy/320/lec/upsampling.pdf>
9. Gjoreski, M., Lutrek, M., Gams, M., & Gjoreski, H. (2017, August 10). Monitoring stress with a wrist device using context. Retrieved September 27, 2020, from <https://www.sciencedirect.com/science/article/pii/S1532046417301855>
10. Mahdiani S, Jeyhani V, Peltokangas M, Vehkaoja A. Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis? Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:5948-51. doi: 10.1109/EMBC.2015.7319746. PMID: 26737646.
11. Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. Retrieved September 26, 2020, from <https://www.biopac.com/wp-content/uploads/EDA-SCR-Analysis.pdf>
12. Pourmohammadi, S., & Maleki, A. (2020, May 05). Stress detection using ECG and EMG signals: A comprehensive study. Retrieved September 27, 2020, from <https://www.sciencedirect.com/science/article/pii/S0169260719320528>
13. https://www.researchgate.net/post/Is_analysis_of_electrodermal_activity_data_possible_with_

a_low_sampling_rate_of_only_2_Hz