Methods & Results Section

# Methods

Data analyses were carried out in R version 3.5.0 (2018-04-23) and Python 2.7. In accordance with previous studies examining the relation between HRV and discrete eating episodes (Friesen et al. 2007; Harthoorn and Dransfield 2007; Ranzenhofer et al. 2016), physiological data collected in the 30-minute period preceding the eating episode were examined. For each participant, the inter-beat interval (IBI) for each self-reported survey was selected, where surveys were either self-reported at the time of an eating episode, or pseudo-randomly triggered in the app. This resulted in an initial total of 487 emotional eating episodes and 3155 control episodes before cleaning the data.

Using the RHRV package (Rodriguez-Linares et al. 2017), IBI data streams were filtered, interpolated, and artifacts were removed, using thresholds and parameters in compliance with normative reported values for heart rate variability features (Shaffer and Ginsberg 2017). Observations that did not meet normative criteria were removed. Consistent with existing research, features of heart rate variability were extracted from both the time and frequency domain to measure both the amount of variability and the amount of signal energy in the 30-minute period (Shaffer and Ginsberg 2017; Rubin et al. 2016).

The following heart rate variability features were extracted from the IBI streams for each observation in the time domain: *SDNN* (Standard Deviation of all filtered inter-beat intervals); *SDANN* (Standard Deviation of inter-beat intervals between all successive heartbeats); *SDNNIX* (Mean of the standard deviations of all the filtered inter-beat intervals for each 5 min segment of the IBI stream); *pNN50* (Percentage of successive inter-beat intervals that differ by more than 50 ms); *SDSD* (Standard Deviation of Successive Differences of inter-beat intervals); *rMSSD* (Root Mean Square of Successive Differences); *IRRR* (length of the interval determined by the first and the third quantile of the inter-beat interval); *MADRR* (Median of the Absolute values of the successive Differences between the inter-beat intervals); *TINN* (Triangular Interpolation of inter-beat interval histogram); and *HRVi* (Heart Rate Variability index) (Rodriguez-Linares et al. 2017).

*SDNN* reflects the power of the components responsible for variability in heart rhythm, while *SDANN*, *SDNNIX*, *pNN50*, *SDSD*, *rMSSD*, *IRRR*, and *MADRR* reflect different statistical aspects of the high frequency variation in heart rhythm. Finally, *TINN* and *HRVi* represent geometric measures calculated from the density distribution of inter-beat intervals (Rodriguez-Linares et al. 2017).

In the frequency domain, features extracted included the mean non-interpolated heart rate (*mean niHR*), the mean interpolated Heart Rate (*mean HR*), and the heart rate values at the start and end of the 30 minute window (*Start HR*, *End HR*). Additional frequency domain features were extracted from the 30 minute streams by applying a Fourier transform to the heart rate signal, resulting in a spectrogram of the heart rate in the low and high frequency ranges. To extract a sufficient amount of information from these spectograms and increase the variance of our data, the spectrogram was split into 5 minute windows, as has been done previously in studies using heart rate variability to predict event-level behavior (Rubin et al. 2016). Mean features were extracted from each of these windows — the mean Low Frequency signal (, ,… ), the mean High Frequency signal (, ,… ), and the mean Low Frequency-High Frequency Ratio (, ,… ).

In order to examine any differences between means of the features, and address the issue of group imbalance affecting statistical power, we conducted between-groups permutation T-tests with each variable in the time and frequency domain (Maindonald and Braun 2015; Good 2013). Then, using the CARET package (Jed Wing et al. 2018), a Support Vector Machine with a polynomial kernel was implemented to predict episodes from controls. In order to address the group imbalance in this stage, all features were standardized, while controls were randomly downsampled and episodes were randomly upsampled using the ROSE package (Lunardon, Menardi, and Torelli 2014). To evaluate model performance, models were tested using 4-fold cross validation, training models on 3/4 of the data and using the remaining data evaluate the model accuracy, specificity, and sensitivity.

# Results

## Missingness

After cleaning the dataset of erroneous surveys (e.g. no usable IBI data; overlapping survey responses; control observations with high self-reported stress), and filtering IBI streams that did not meet filtering quality criteria for preprocessing, the data consisted of 285 emotional eating episodes and 1753 controls. Of these clean observations, HRV features were extracted from the IBI streams, although even with filtering, many IBI streams were too sparse to compute HRV features. Figures 1 and 2 illustrate the completeness of data in the extracted time and frequency domain features, with tables 1 and 2 showing the final number of observations used for analysis. Much of the data missing from the time domain is due to the fact that these features derived from SDNN and are dependent on having sufficient data in this feature.

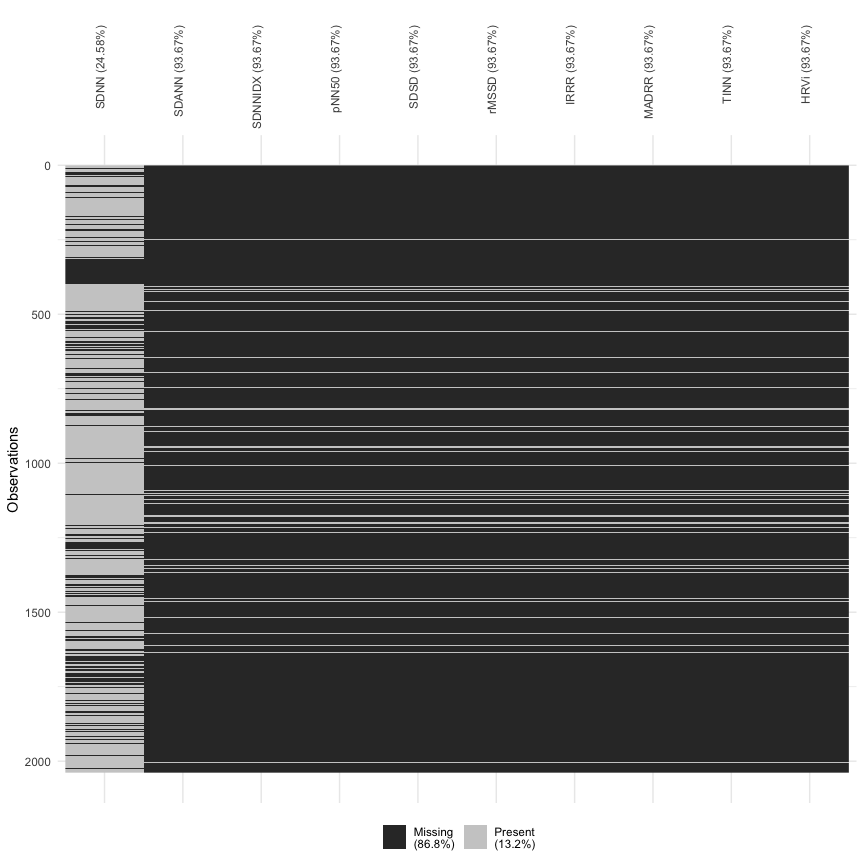


Figure 1 Completeness of Features in the Time Domain

Table 1 Number of Time Domain Observations

|  |  |  |
| --- | --- | --- |
| Y | Control | Episode |
| SDNN | 1318 | 219 |
| SDANN | 110 | 19 |
| SDNNIDX | 110 | 19 |
| pNN50 | 110 | 19 |
| SDSD | 110 | 19 |
| rMSSD | 110 | 19 |
| IRRR | 110 | 19 |
| MADRR | 110 | 19 |
| TINN | 110 | 19 |
| HRVi | 110 | 19 |

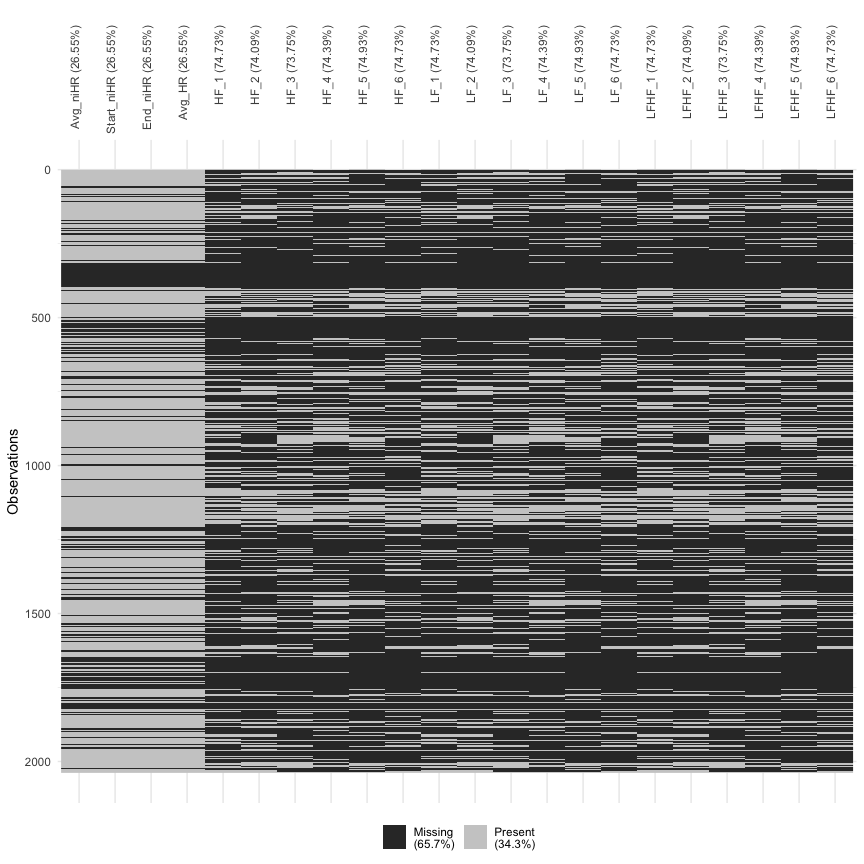


Figure 2 Completeness of Features in the Frequency Domain

Table 2 Number of Frequency Domain Observations

|  |  |  |
| --- | --- | --- |
| Y | Control | Episode |
| **Heart Rate** | | |
| Avg\_niHR | 1282 | 215 |
| Start\_niHR | 1282 | 215 |
| End\_niHR | 1282 | 215 |
| Avg\_HR | 1282 | 215 |
| **High Frequency** | | |
| HF\_1 | 433 | 82 |
| HF\_2 | 451 | 77 |
| HF\_3 | 456 | 79 |
| HF\_4 | 440 | 82 |
| HF\_5 | 445 | 66 |
| HF\_6 | 444 | 71 |
| **Low Frequency** | | |
| LF\_1 | 433 | 82 |
| LF\_2 | 451 | 77 |
| LF\_3 | 456 | 79 |
| LF\_4 | 440 | 82 |
| LF\_5 | 445 | 66 |
| LF\_6 | 444 | 71 |
| **LF-HF Ratio** | | |
| LFHF\_1 | 433 | 82 |
| LFHF\_2 | 451 | 77 |
| LFHF\_3 | 456 | 79 |
| LFHF\_4 | 440 | 82 |
| LFHF\_5 | 445 | 66 |
| LFHF\_6 | 444 | 71 |

## Comparison of Means

The mean values for each time-domain features is shown in Table 3 below:

Table 3 Summary of Features in the Time Domain

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | ***n*** | **Mean** | **SD** |
| HRVi | 129 | 7.171186 | 1.992292 |
| IRRR | 129 | 65.288888 | 28.536286 |
| MADRR | 129 | 26.043073 | 10.932954 |
| pNN50 | 129 | 19.830466 | 10.590041 |
| rMSSD | 129 | 48.539227 | 13.137957 |
| SDANN | 129 | 15.440406 | 15.451301 |
| SDNN | 1537 | 69.429465 | 24.835250 |
| SDNNIDX | 129 | 43.869882 | 12.781756 |
| SDSD | 129 | 48.557984 | 13.143443 |
| TINN | 129 | 112.049788 | 31.129567 |

Using a Two *t* permutation test run between episodes and controls, none of the time-domain variables were found to be significantly different, shown in Table 4.

Table 4 Between-Groups Permutation *t*-test of Time Domain Features

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | | ***n*** | | **SD** | |  |
| **Variable** | **Control** | **Episode** | **Control** | **Episode** | **Control** | **Episode** | ***p*** |
| HRVi | 7.20 | 6.98 | 110 | 19 | 2.12 | 1.04 | 0.66 |
| IRRR | 66.05 | 60.86 | 110 | 19 | 30.50 | 11.52 | 0.47 |
| MADRR | 26.14 | 25.49 | 110 | 19 | 11.22 | 9.33 | 0.73 |
| pNN50 | 19.99 | 18.93 | 110 | 19 | 10.93 | 8.54 | 0.69 |
| rMSSD | 48.95 | 46.17 | 110 | 19 | 13.61 | 9.95 | 0.39 |
| SDANN | 15.62 | 14.40 | 110 | 19 | 16.33 | 9.02 | 0.74 |
| SDNN | 69.69 | 67.87 | 1318 | 219 | 25.23 | 22.32 | 0.31 |
| SDNNIDX | 44.04 | 42.86 | 110 | 19 | 13.45 | 8.05 | 0.70 |
| SDSD | 48.97 | 46.19 | 110 | 19 | 13.61 | 9.96 | 0.39 |
| TINN | 112.55 | 109.14 | 110 | 19 | 33.06 | 16.20 | 0.66 |

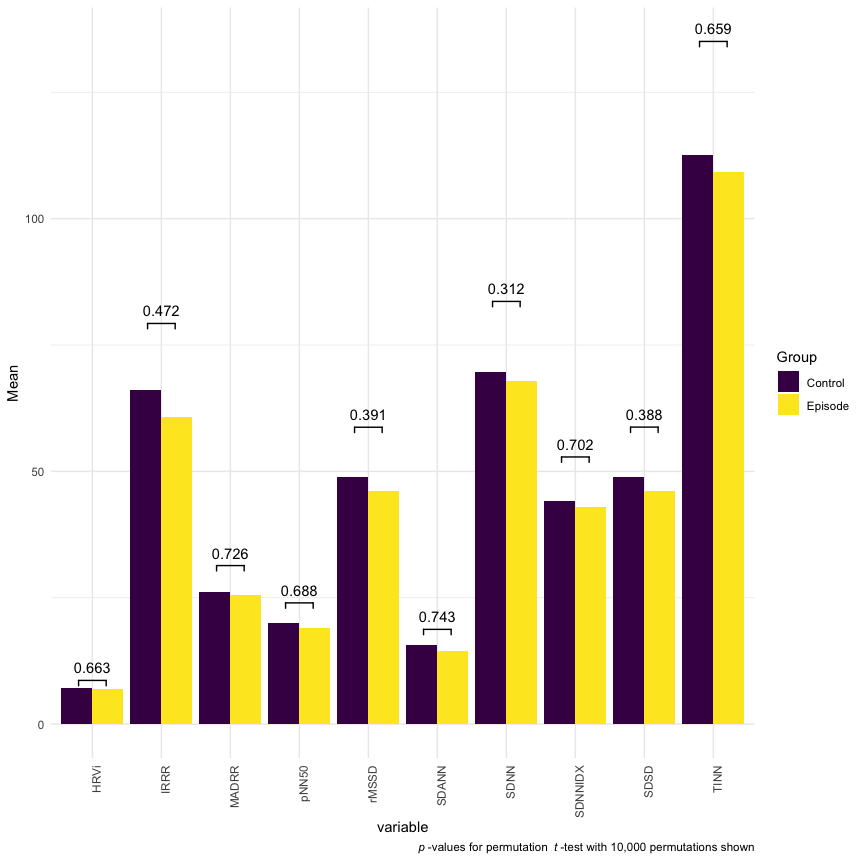


Figure 3 Between-Groups Permutation t-test of Time Domain Features

A similar approach was taken to analyze the frequency domain. Table 5 below outlines the mean values of features:

Table 5 Summary of Features in the Frequency Domain

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | ***n*** | **Mean** | **SD** |
| **Heart Rate** | | | |
| Avg\_HR | 1497 | 79.50 | 12.13 |
| Avg\_niHR | 1497 | 80.26 | 12.20 |
| Start\_niHR | 1497 | 83.58 | 17.76 |
| End\_niHR | 1497 | 81.34 | 15.69 |
| **High Frequency** | | | |
| HF\_1 | 515 | 282.34 | 382.80 |
| HF\_2 | 528 | 271.35 | 351.14 |
| HF\_3 | 535 | 262.68 | 291.35 |
| HF\_4 | 522 | 248.05 | 289.56 |
| HF\_5 | 511 | 254.59 | 317.44 |
| HF\_6 | 515 | 295.68 | 432.77 |
| **Low Frequency** | | | |
| LF\_1 | 515 | 320.98 | 254.62 |
| LF\_2 | 528 | 319.81 | 246.48 |
| LF\_3 | 535 | 323.36 | 247.22 |
| LF\_4 | 522 | 314.75 | 238.92 |
| LF\_5 | 511 | 334.37 | 253.86 |
| LF\_6 | 515 | 345.29 | 269.47 |
| **LF-HF Ratio** | | | |
| LFHF\_1 | 515 | 4.69 | 3.42 |
| LFHF\_2 | 528 | 4.68 | 3.53 |
| LFHF\_3 | 535 | 4.59 | 3.46 |
| LFHF\_4 | 522 | 4.83 | 3.55 |
| LFHF\_5 | 511 | 4.98 | 3.55 |
| LFHF\_6 | 515 | 4.85 | 3.42 |

Using the same Two-*t* permutation approach, we found signficant differences in means for Average Heart Rate, Average non-interpolated Heart Rate, Ending non-interpolated Heart Rate, and the LF-HF ratio in the third window, shown in Table 6

Table 6 Between-Groups Permutation *t*-test of Frequency Domain Features

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | | ***n*** | | **SD** | |  |
| **Variable** | **Control** | **Episode** | **Control** | **Episode** | **Control** | **Episode** | ***p*** |
| **Heart Rate** | | | | | | | |
| Avg\_HR | 79.22 | 81.20 | 1282 | 215 | 12.25 | 11.30 | 0.03 |
| Avg\_niHR | 80.01 | 81.76 | 1282 | 215 | 12.31 | 11.41 | 0.05 |
| Start\_niHR | 83.56 | 83.70 | 1282 | 215 | 18.21 | 14.85 | 0.92 |
| End\_niHR | 80.77 | 84.71 | 1282 | 215 | 15.55 | 16.12 | 0.00 |
| **High Frequency** | | | | | | | |
| HF\_1 | 272.59 | 333.79 | 433 | 82 | 329.02 | 591.06 | 0.18 |
| HF\_2 | 263.68 | 316.23 | 451 | 77 | 305.91 | 546.34 | 0.22 |
| HF\_3 | 257.48 | 292.74 | 456 | 79 | 257.94 | 438.11 | 0.31 |
| HF\_4 | 244.61 | 266.51 | 440 | 82 | 259.52 | 416.97 | 0.55 |
| HF\_5 | 262.25 | 202.95 | 445 | 66 | 335.09 | 143.31 | 0.14 |
| HF\_6 | 305.57 | 233.80 | 444 | 71 | 460.86 | 163.01 | 0.18 |
| **Low Frequency** | | | | | | | |
| LF\_1 | 322.90 | 310.85 | 433 | 82 | 255.19 | 252.93 | 0.70 |
| LF\_2 | 315.46 | 345.28 | 451 | 77 | 245.66 | 251.30 | 0.33 |
| LF\_3 | 326.12 | 307.45 | 456 | 79 | 253.17 | 210.27 | 0.53 |
| LF\_4 | 309.79 | 341.38 | 440 | 82 | 235.44 | 256.66 | 0.27 |
| LF\_5 | 330.80 | 358.43 | 445 | 66 | 252.93 | 260.77 | 0.41 |
| LF\_6 | 347.62 | 330.70 | 444 | 71 | 273.40 | 244.74 | 0.63 |
| **LF-HF Ratio** | | | | | | | |
| LFHF\_1 | 4.72 | 4.57 | 433 | 82 | 3.46 | 3.20 | 0.71 |
| LFHF\_2 | 4.68 | 4.67 | 451 | 77 | 3.50 | 3.72 | 0.97 |
| LFHF\_3 | 4.72 | 3.86 | 456 | 79 | 3.49 | 3.19 | 0.04 |
| LFHF\_4 | 4.78 | 5.09 | 440 | 82 | 3.54 | 3.60 | 0.46 |
| LFHF\_5 | 4.98 | 5.02 | 445 | 66 | 3.52 | 3.74 | 0.94 |
| LFHF\_6 | 4.80 | 5.14 | 444 | 71 | 3.43 | 3.35 | 0.44 |

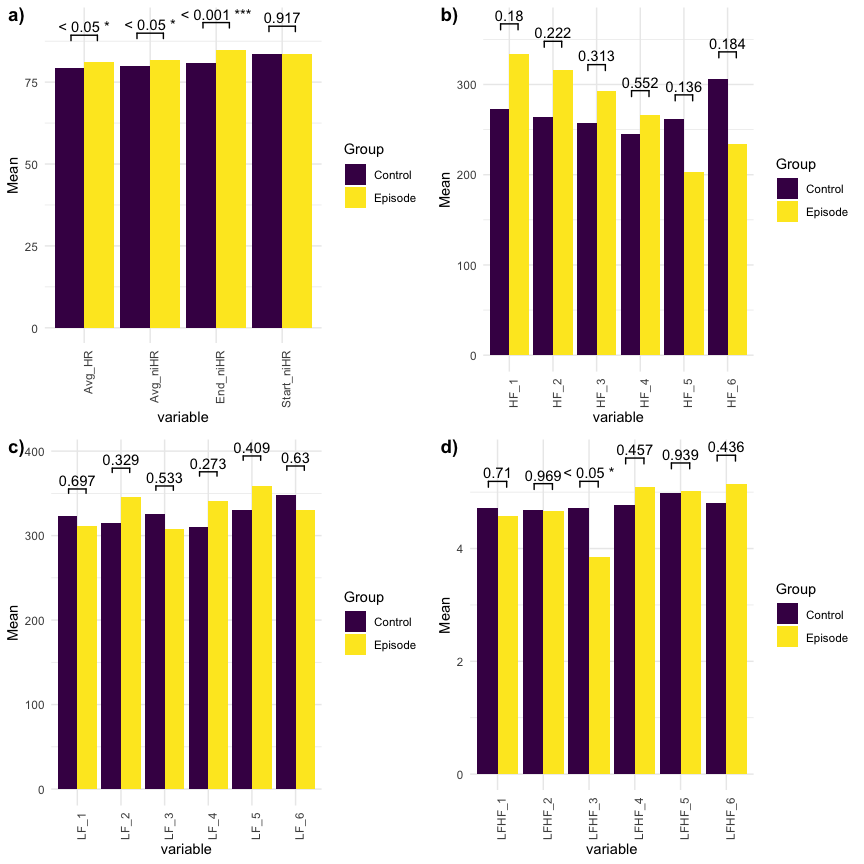


Figure 4 Between-Groups Permutation t-test of Frequency Domain Features: a) Heart Rate Features; b) High Frequency Features; c) Low Frequency Features d) Low Frequency-High Frequency Ratio Features

## Machine Learning

Machine learning models were implemented using time domain and frequency domain features separately. In order to run machine learning prediction of emotional eating episodes, observations with complete data for all features in the time or frequency domain were used, data within each participant were scaled and centered, and observations were randomly upsampled (from episodes) or downsampled (from controls) to balance the number of observations and controls. This resulted in 126 training observations, and 32 test observations in the time domain models, with 95 training observations, and 24 test observations in the frequency domain models. SVM models were fit iteratively using 3/4’s of the data, and evaluated on the remaining 1/4. Figure 5 show the mean accuracy, sensitivity, and specificity of the SVM across these four folds.

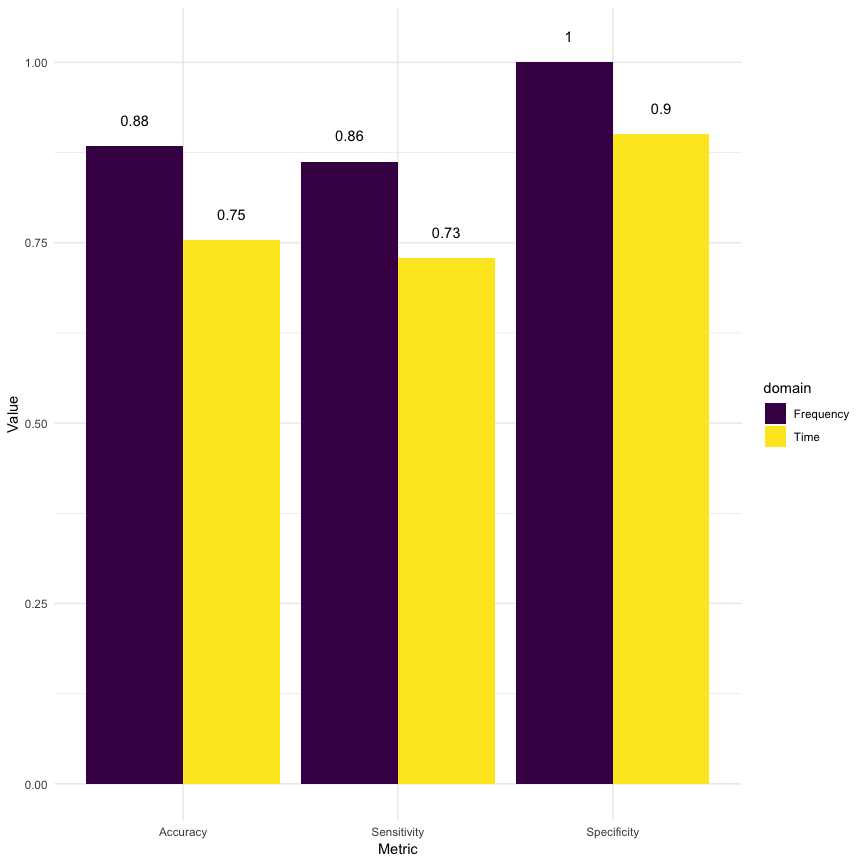


Figure 5 Machine Learning Evaluation Metrics

To interpret the feature importance in each of these models, ROC curve analysis was conducted on each feature on the model as recommended by Khun Jed Wing et al. (2018). This approach provides a scale-invariant metric that illustrates the impact of a feature on the prediction based on the Area Under the Curve (AUC) of the ROC of a classifier with only the single feature. Figures 6 and 7 show the results of this analysis.

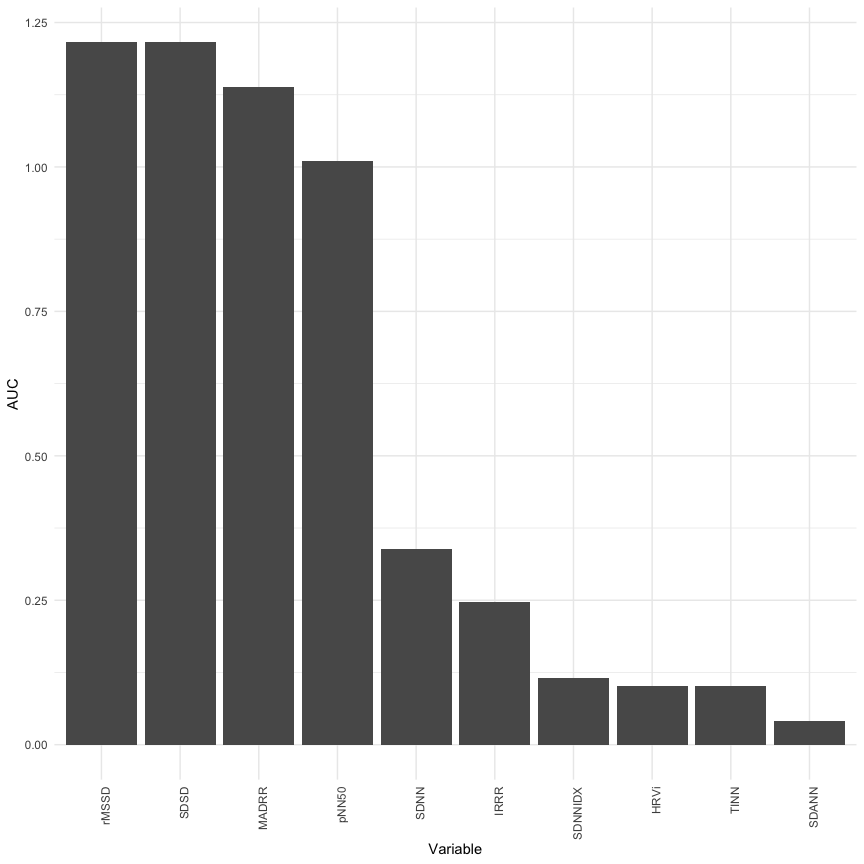


Figure 6 Feature Importance of Features in the Time Domain

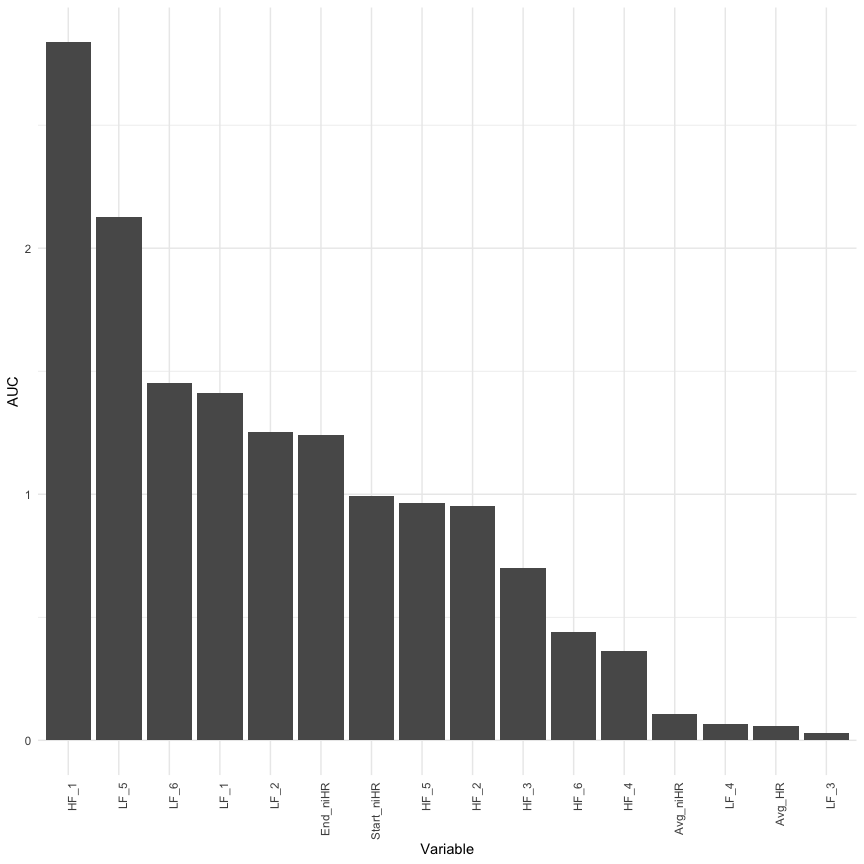


Figure 7 Feature Importance of Features in the Frequency Domain

# Discussion

A major hurdle to producing interpretable results is the sanity of data obtained from the devices. Despite best efforts, figures 1 and 2 illustrate how sparse the data can be and how this affects its overall usefulness in running analyses.The causes for this data sparsity are unclear, and could include reasons such as improper usage or low signal while recording data. The result of this challenge is that derived features are incalculable; for example, time domain feature calculations are strongly dependent on a sliding window that must have within it at least some number of consecutive heart beats, or completeness and consecutiveness of *SDNN*; frequency domain variables are dependent on Fourier transforms which are sensitive to small and variations in wave frequencies. Nevertheless, by following proper data sanity checks and procedures provided in analyses packages and in the literature, it’s still possible to extract useful information from these data.

Although no features in the time domain were significantly different between episode and control observations when tested, it’s worth mentioning that the machine learning model was still able to achieve decent classification accuracy, sensitivity, and specificity using these features. The most impactful features in this model were *rMSSD* (Root Mean Square of Successive Differences); *SDSD* (Standard Deviation of Successive Differences of inter-beat intervals); and *MADRR* (Median of the Absolute values of the successive Differences between the inter-beat intervals). However, the frequency domain achieved superior classification accuracy in comparison to the time domain, with mean accuracy of 88%, sensitivity of 86%, and specificity of 90%. For this model, the greatest variable importance was for , , , , and . This may reflect a tendency of heart rate to change most noticeably in the low frequency power band, in the moments leading up to an eating episode.

Friesen, C. A., Z. Lin, J. V. Schurman, L. Andre, and R. W. Mc Callum. 2007. “Autonomic Nervous System Response to a Solid Meal and Water Loading in Healthy Children: Its Relation to Gastric Myoelectrical Activity.” *Neurogastroenterology & Motility* 19 (5): 376–82. <https://doi.org/10.1111/j.1365-2982.2007.00906.x>.

Good, Phillip. 2013. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Science & Business Media.

Harthoorn, Lucien F., and Eric Dransfield. 2007. “Periprandial Changes of the Sympathetic–Parasympathetic Balance Related to Perceived Satiety in Humans.” *European Journal of Applied Physiology* 102 (5): 601–8. <https://doi.org/10.1007/s00421-007-0622-5>.

Jed Wing, Max Kuhn. Contributions from, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, et al. 2018. *Caret: Classification and Regression Training*. <https://CRAN.R-project.org/package=caret>.

Lunardon, Nicola, Giovanna Menardi, and Nicola Torelli. 2014. *ROSE: Random over-Sampling Examples*. <https://CRAN.R-project.org/package=ROSE>.

Maindonald, John H., and W. John Braun. 2015. *DAAG: Data Analysis and Graphics Data and Functions*. <https://CRAN.R-project.org/package=DAAG>.

Ranzenhofer, Lisa M, Scott G Engel, Ross D Crosby, Mark Haigney, Micheline Anderson, Jeanne M McCaffery, and Marian Tanofsky-Kraff. 2016. “Real-Time Assessment of Heart Rate Variability and Loss of Control Eating in Adolescent Girls: A Pilot Study.” *International Journal of Eating Disorders* 49 (2): 197–201.

Rodriguez-Linares, Leandro, Xose Vila, Maria Jose Lado, Arturo Mendez, Abraham Otero, and Constantino Antonio Garcia. 2017. *RHRV: Heart Rate Variability Analysis of Ecg Data*. <https://CRAN.R-project.org/package=RHRV>.

Rubin, Jonathan, Rui Abreu, Shane Ahern, Hoda Eldardiry, and Daniel G. Bobrow. 2016. “Time, Frequency &#38; Complexity Analysis for Recognizing Panic States from Physiologic Time-Series.” In *Proceedings of the 10th Eai International Conference on Pervasive Computing Technologies for Healthcare*, 81–88. PervasiveHealth ’16. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics; Telecommunications Engineering). <http://dl.acm.org/citation.cfm?id=3021319.3021332>.

Shaffer, Fred, and J. P. Ginsberg. 2017. “An Overview of Heart Rate Variability Metrics and Norms.” *Frontiers in Public Health* 5. <https://doi.org/10.3389/fpubh.2017.00258>.