Neonatal Sepsis prediction in the NICU using supervised learning techniques

Nadia Aly*, Evan Dienstman

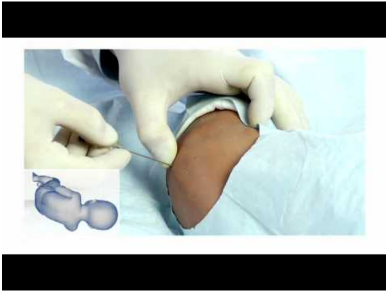
John Delos, PhD., Daniel McGibney, PhD., Anh Ninh, PhD.

[1] College of William and Mary, Department of Mathematics

March 24, 2017

Motivation

- Bacterial Sepsis is the leading cause of mortality in U.S. NICUs
- Current diagnostics are invasive
- If early signs or risk factors are missed, mortality increases and residual neurologic damage is possible



What is Sepsis ?

- Inflammatory response caused by the body fighting off an infection
- Early treatment with antibiotics saves lives
- Premature babies are highly susceptible due to a weak immune system

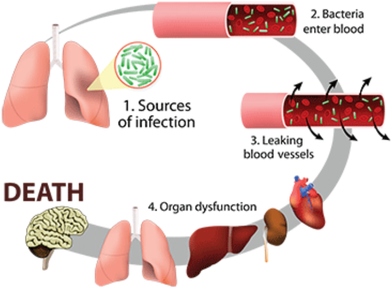


FIGURE – Diagram of Sepsis Infection, NYSUT Health Care Professionals Council

Current Methodology

Data is from a Randomized Clinical Trial from 8 Hospitals : UVA, Wake Forest, UAl, Vanderbilt, Umiami, Greenville SC, Palmer, Penn State.

Experiment	2989 VLBW	Mortality Rate	P-Value	1513 ELBW	Mortality Rate	P-Value
Control	$\frac{152 deaths}{1489}$	10.2%	0.04	$\frac{133 deaths}{757}$	17.6%	0.02
Sample, Display HERO score	$\frac{122 deaths}{1500}$	8.1%	0.04	$\frac{100 deaths}{756}$	13.2%	0.02

TABLE – Results from Randomized Trial

1. Medical Predictive Science Corporation has developed and is marketing a system for Heart Rate Characteristics monitoring in the NICU.
2. A computer beside each NICU bed continuously collects ECG data, extracts times of R peaks, tracks RR intervals, and provides the following Heart Rate Observation (HeRO).

HeRo Score Monitor

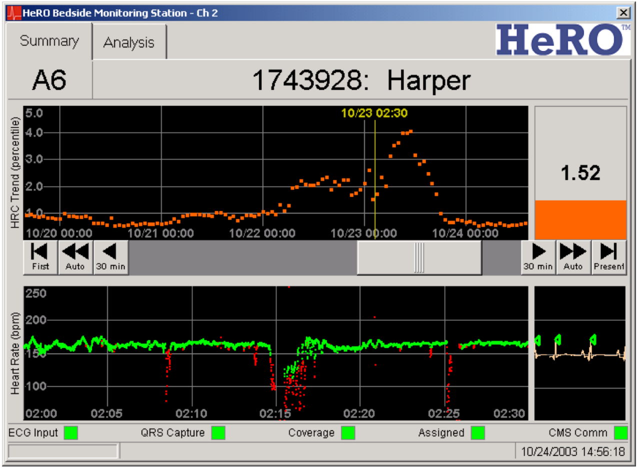


FIGURE – HeRo Monitor Display : Blood culture on 10/23 at 8 AM. Infant was having feeding intolerance as a clinical sign. The heart rate pattern shows decreased variability and decelerations.

Organism Distribution Invading Organisms

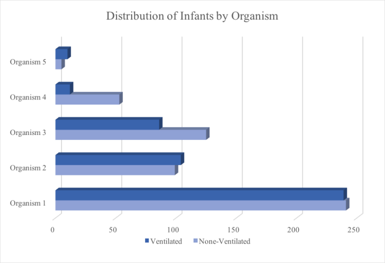


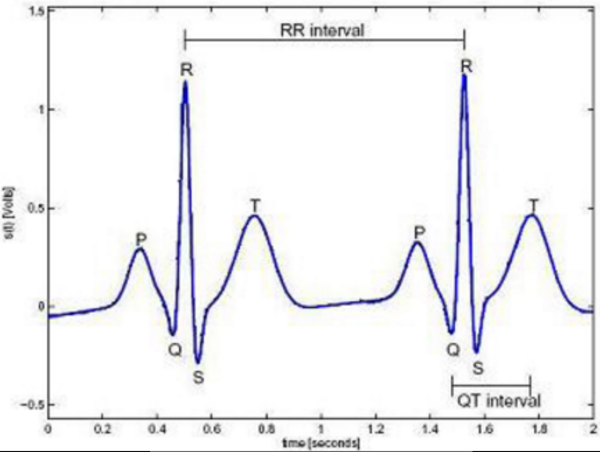
FIGURE – Histogram of Invading Organisms

- Organism 1 :Coagulase-negative Staphylococcal
- Organism 2 :Gram-positive
- Organism 3 :Gram-negative
- Organism 4 :Fungal
- Organism 5 :Other

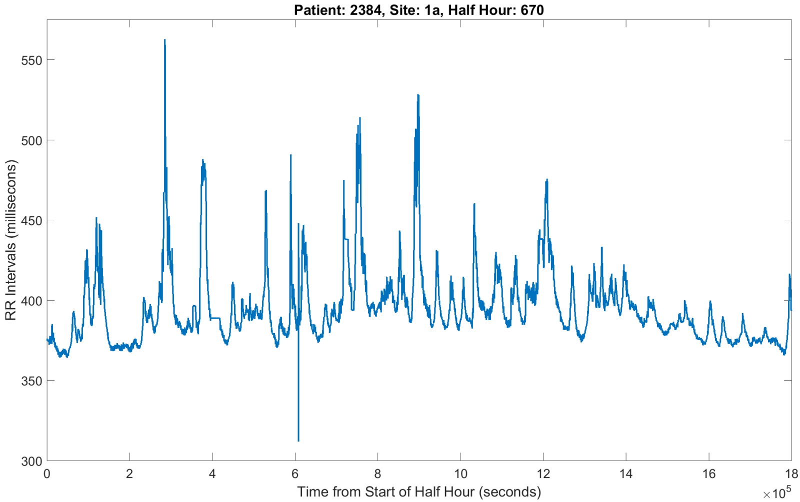
What does the raw data look like ?- RR Interval

rr_intervals = [420.25, 421.75, 420.25, 418.25, 422.00, 423.75 ...]

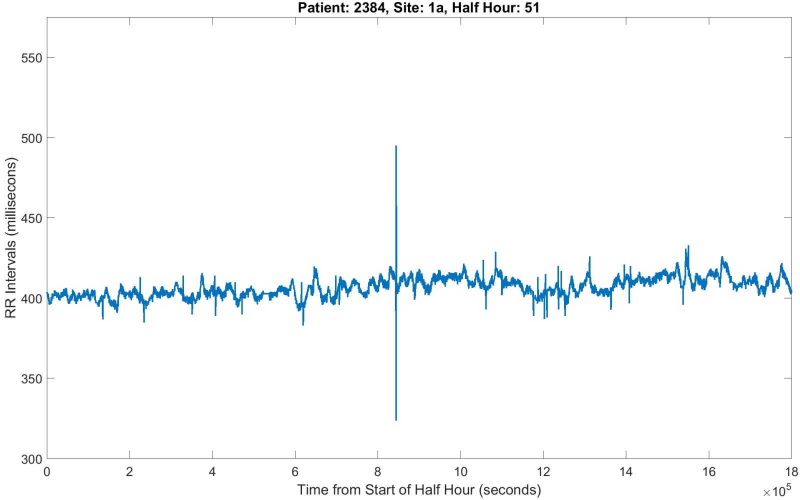
tt_intervals = [420.25, 842.00, 1262.3, 370470, 370880, ...]



What does the raw data look like ?



Does heart rate give warning of illness ?



Does heart rate give warning of illness ?

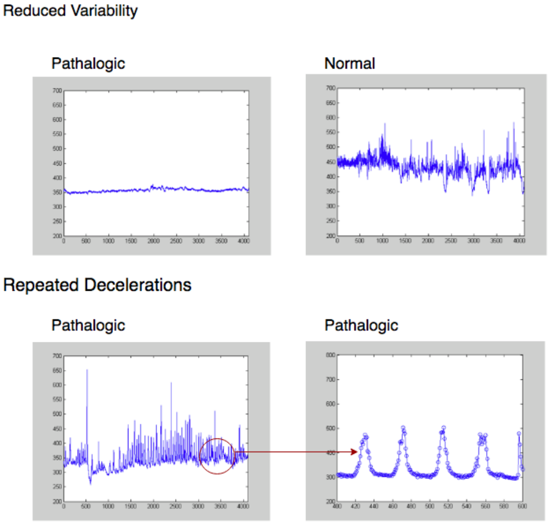


FIGURE – Heart Rate Variability and Decelerations

Does heart rate give warning of illness ?

1. Standard Deviation and Sample Entropy : Variability in the signal.
2. Sample Asymmetry : Prevalence of decelerations over accelerations implies a skew, or asymmetry, in the data which we can detect statistically.

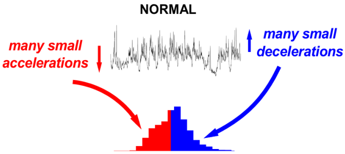


FIGURE – Heart Rate Variability and Decelerations

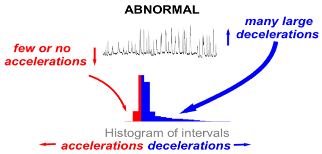


FIGURE – Heart Rate Variability and Decelerations

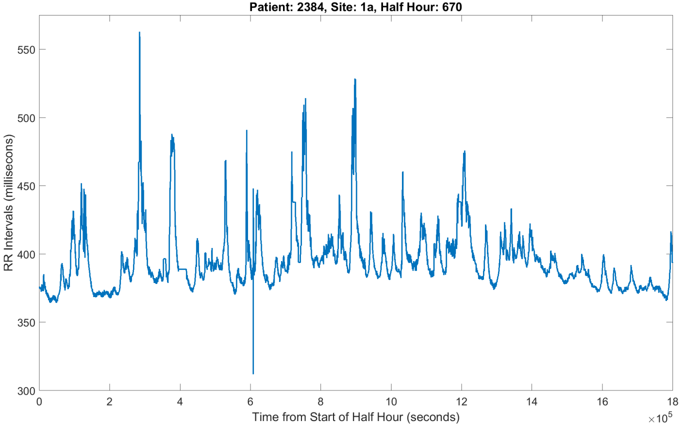
Heart Rate Characteristics

- Variance
- Sample Asymmetry
- Sample Entropy
- Decelerations

Heart Rate Characteristics : Variance

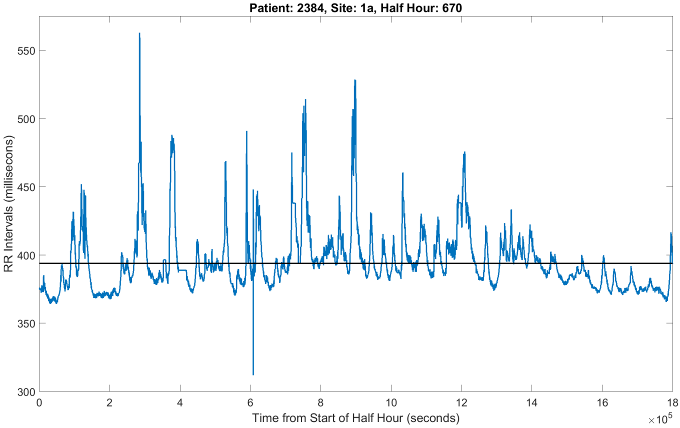
- Variance
- Sample Asymmetry
- Sample Entropy
- Decelerations

Heart Rate Characteristics : Variance

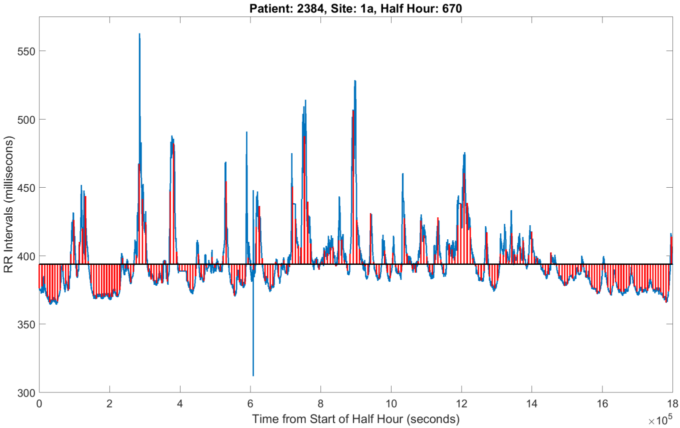


- $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}$

Heart Rate Characteristics : Variance



Heart Rate Characteristics : Variance

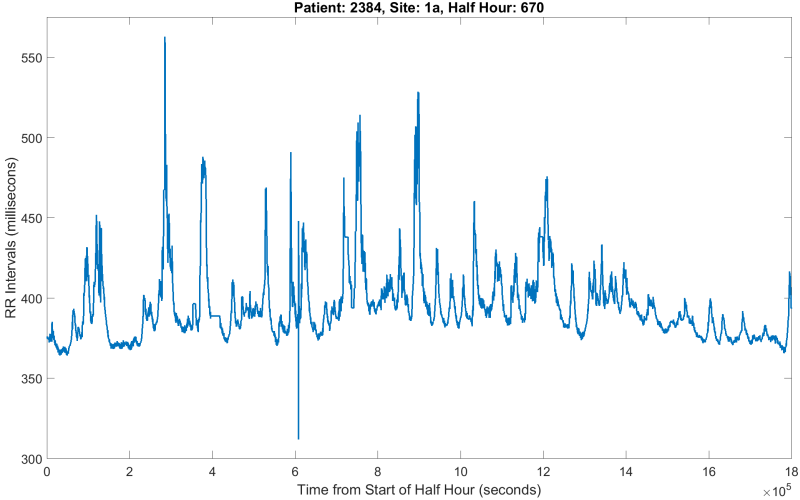


- $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}$

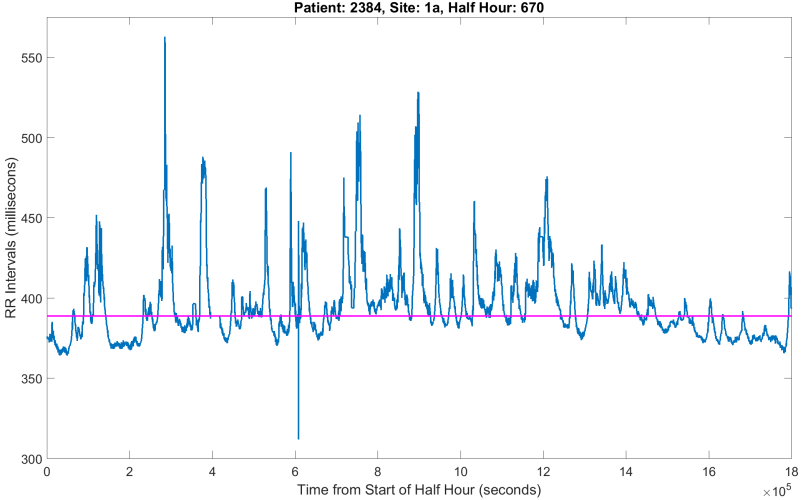
Heart Rate Characteristics : Sample Asymmetry

- Variance
- Sample Asymmetry
- Sample Entropy
- Decelerations

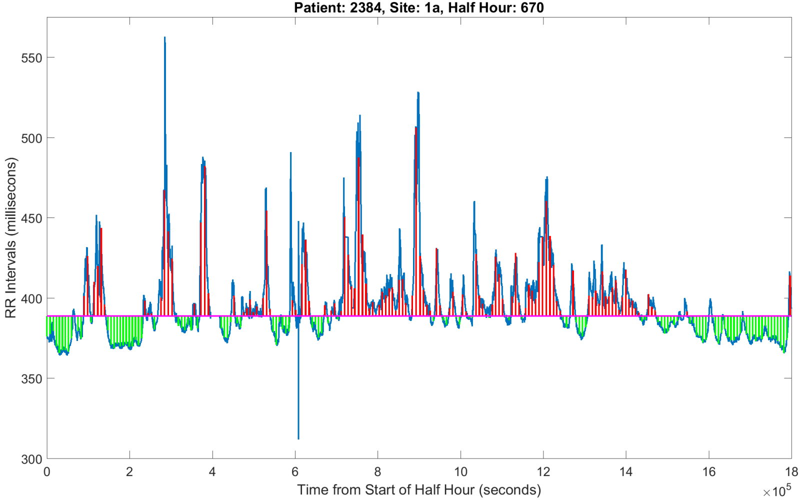
Heart Rate Characteristics : Sample Asymmetry



Heart Rate Characteristics : Sample Asymmetry



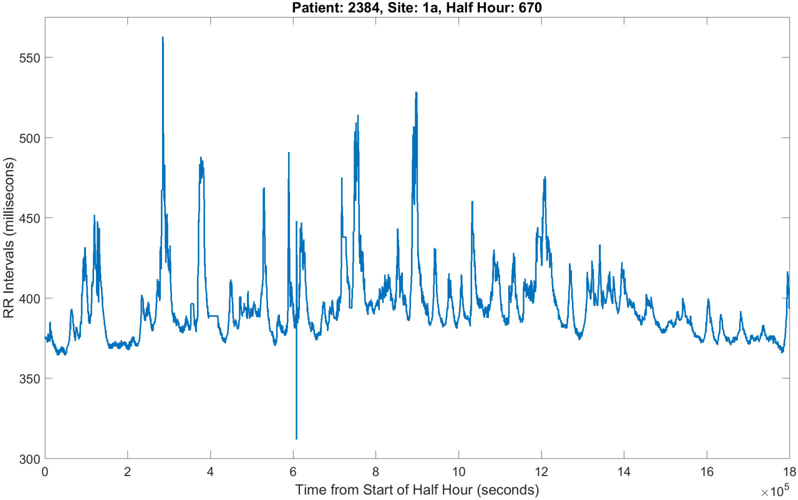
Heart Rate Characteristics : Sample Asymmetry



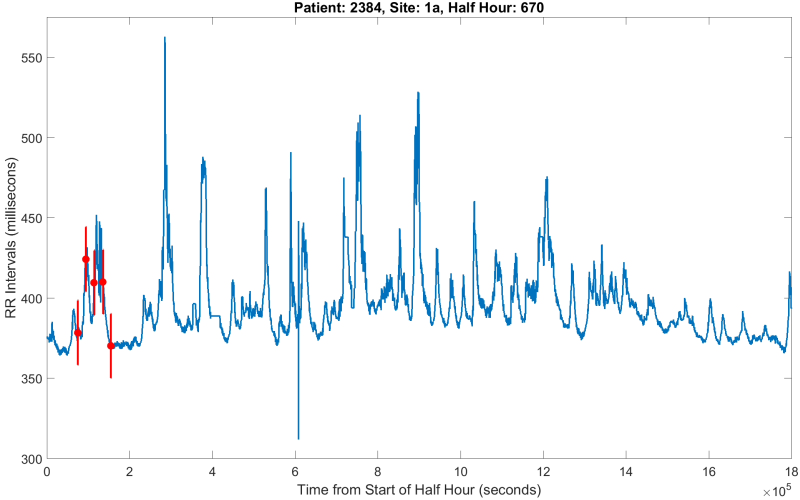
Heart Rate Characteristics : Sample Entropy

- Variance
- Sample Asymmetry
- Sample Entropy
- Decelerations

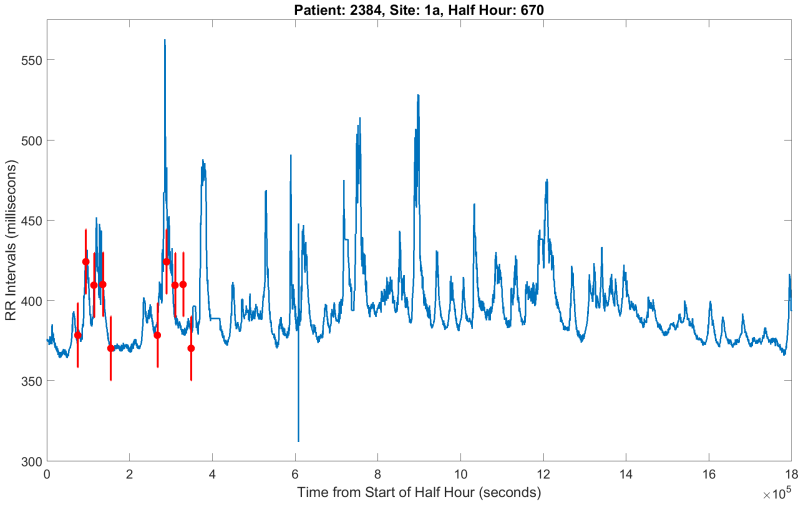
Heart Rate Characteristics : Sample Entropy



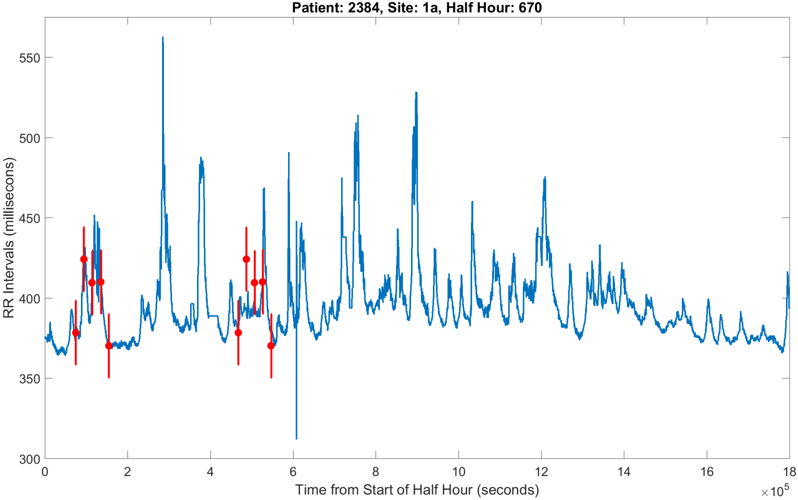
Heart Rate Characteristics : Sample Entropy



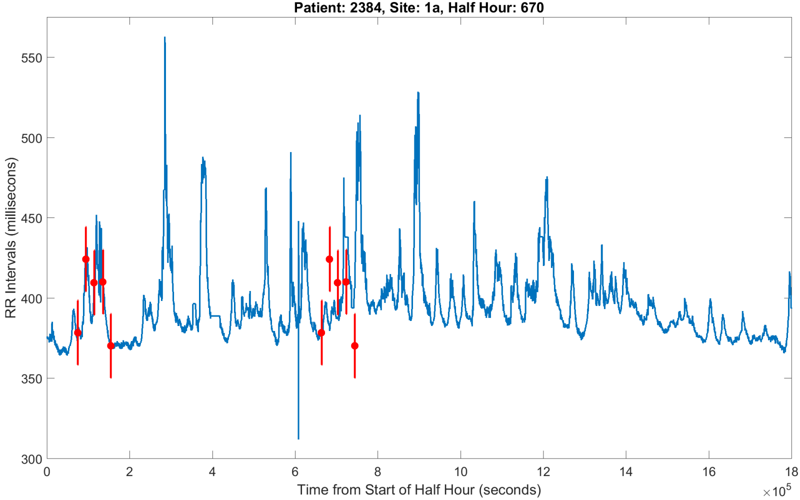
Heart Rate Characteristics : Sample Entropy



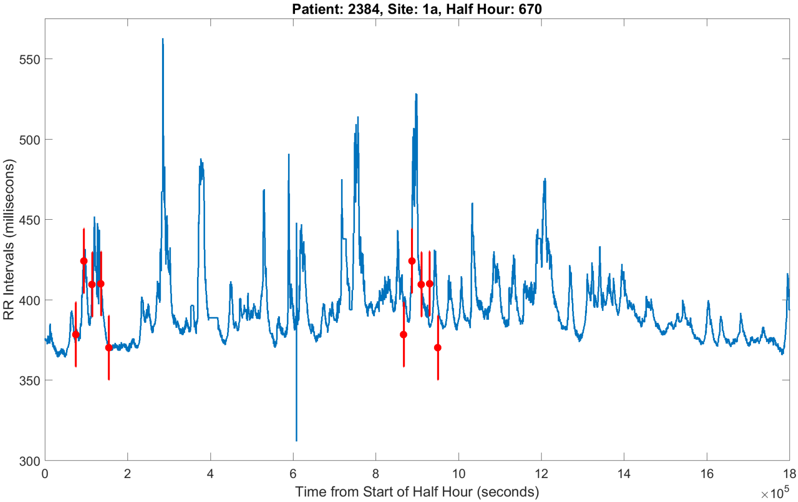
Heart Rate Characteristics : Sample Entropy



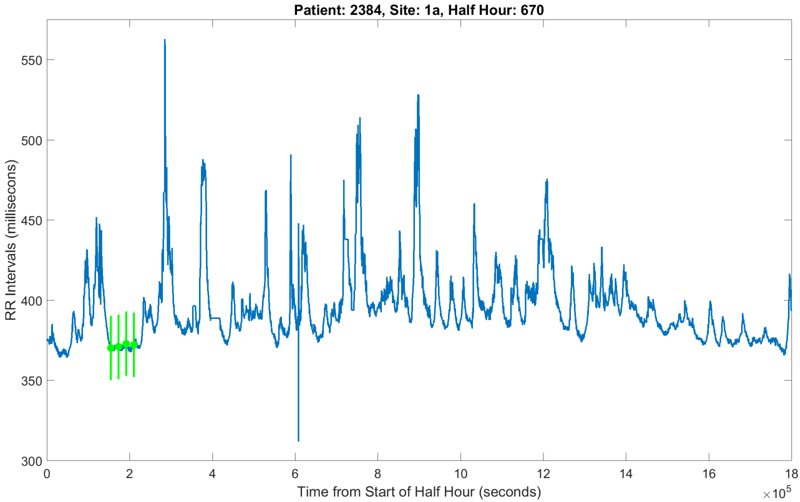
Heart Rate Characteristics : Sample Entropy



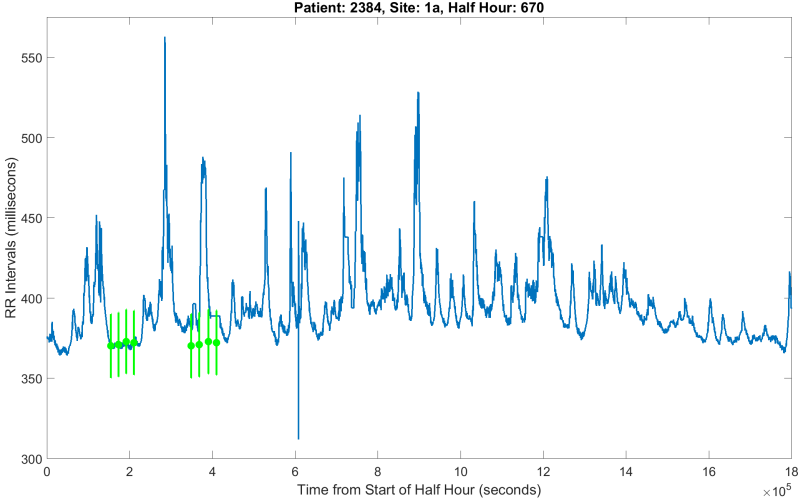
Heart Rate Characteristics : Sample Entropy



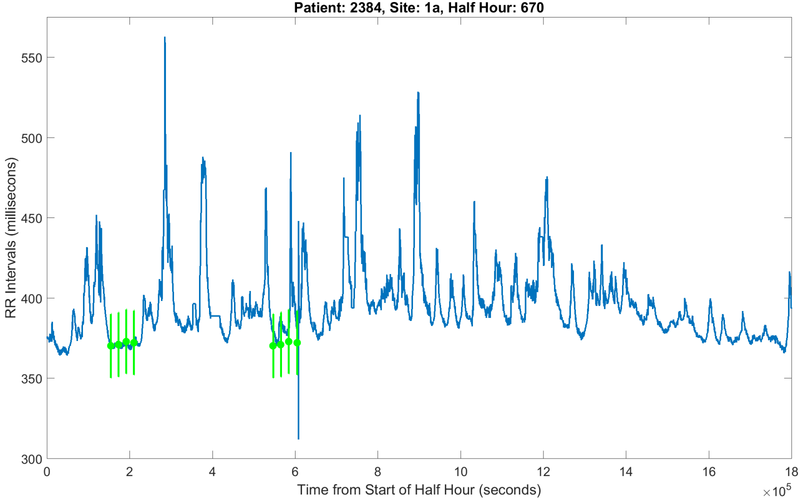
Heart Rate Characteristics : Sample Entropy



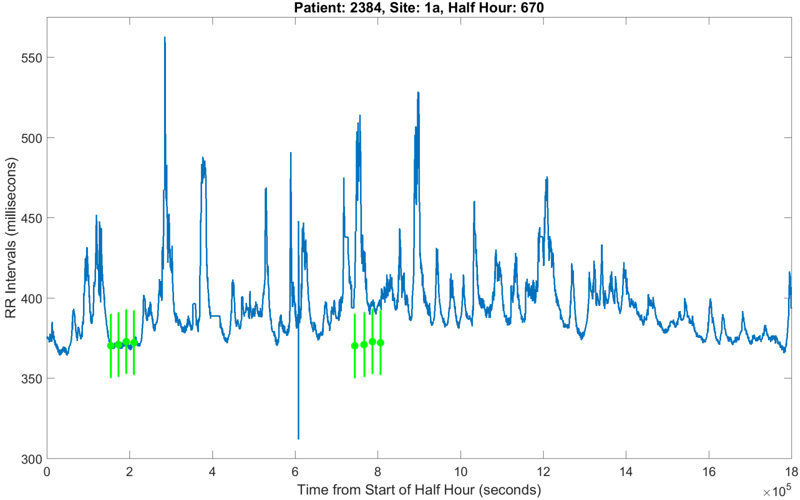
Heart Rate Characteristics : Sample Entropy



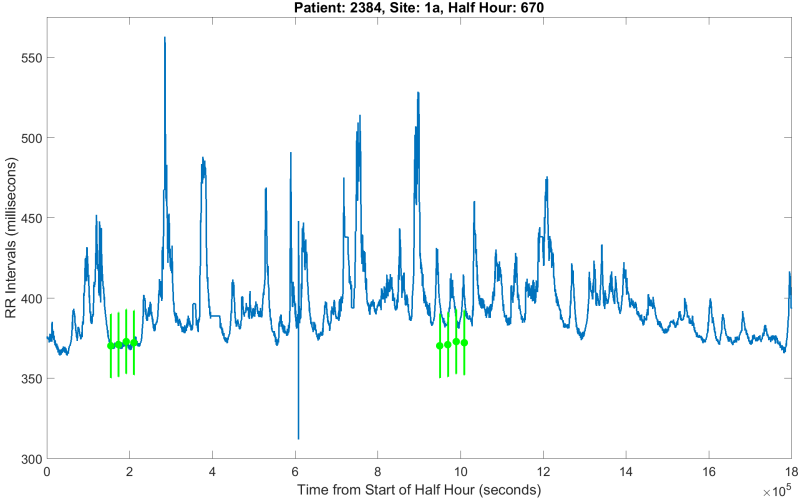
Heart Rate Characteristics : Sample Entropy



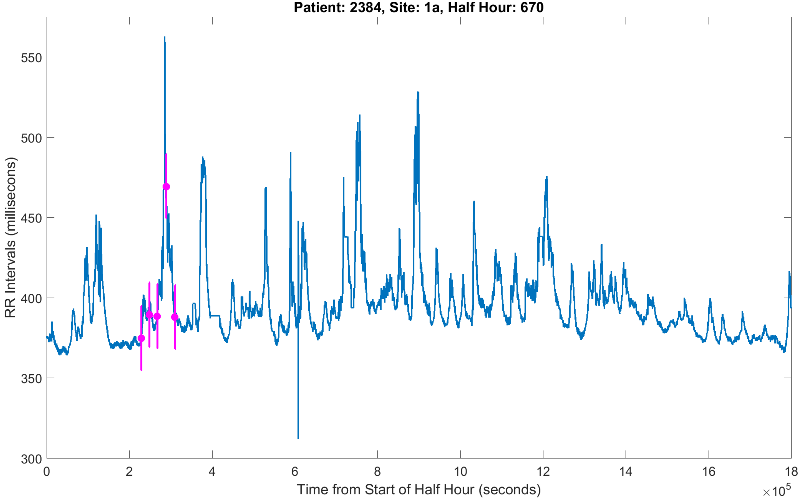
Heart Rate Characteristics : Sample Entropy



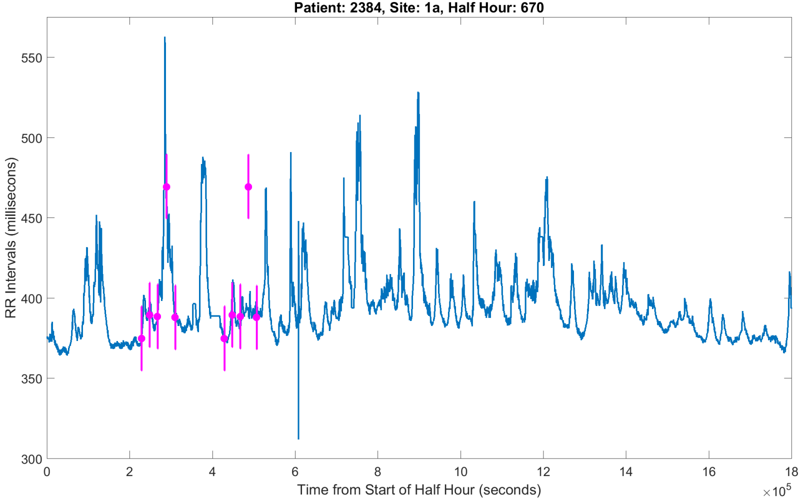
Heart Rate Characteristics : Sample Entropy



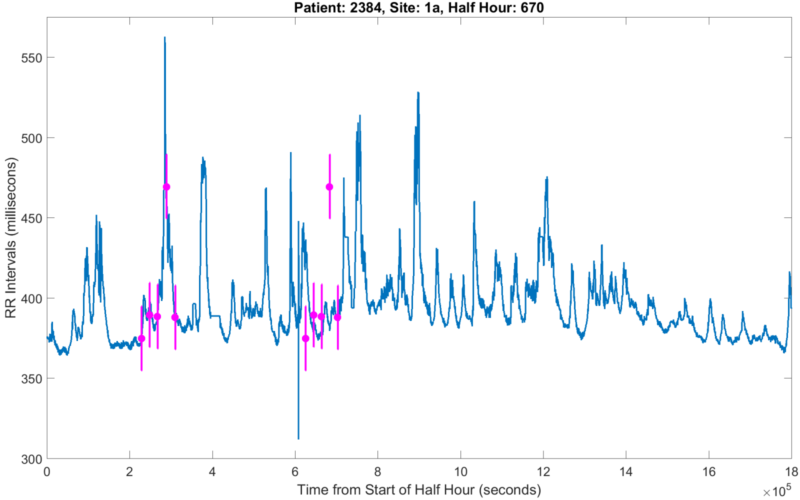
Heart Rate Characteristics : Sample Entropy



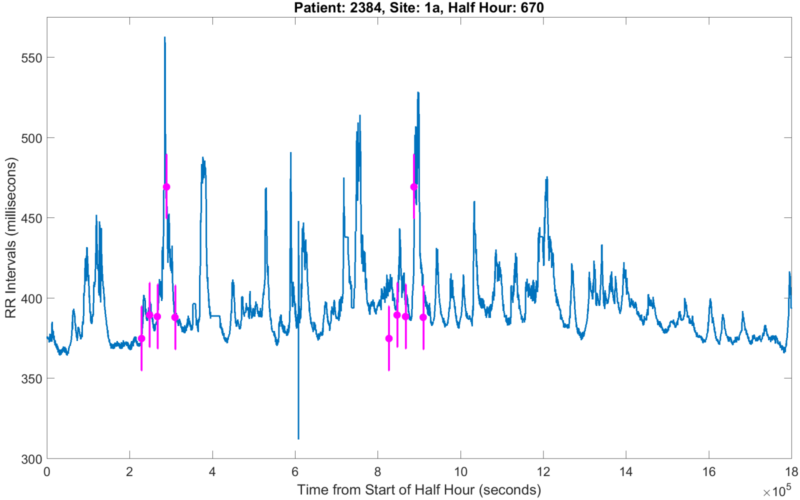
Heart Rate Characteristics : Sample Entropy



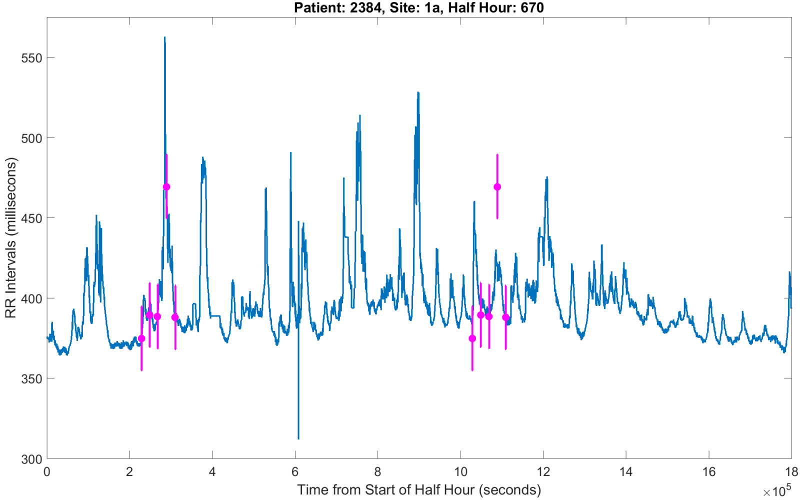
Heart Rate Characteristics : Sample Entropy



Heart Rate Characteristics : Sample Entropy



Heart Rate Characteristics : Sample Entropy

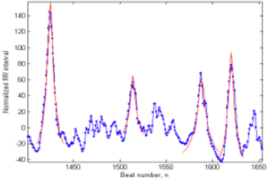


Heart Rate Characteristics : Decelerations

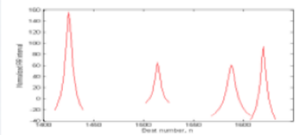
- Variance
- Sample Asymmetry
- Sample Entropy
- Decelerations

Heart Rate Characteristics : Decelerations, Pattern Recognition

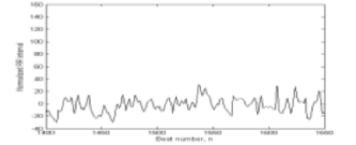
Decompose the signal into decelerations and “background variability :



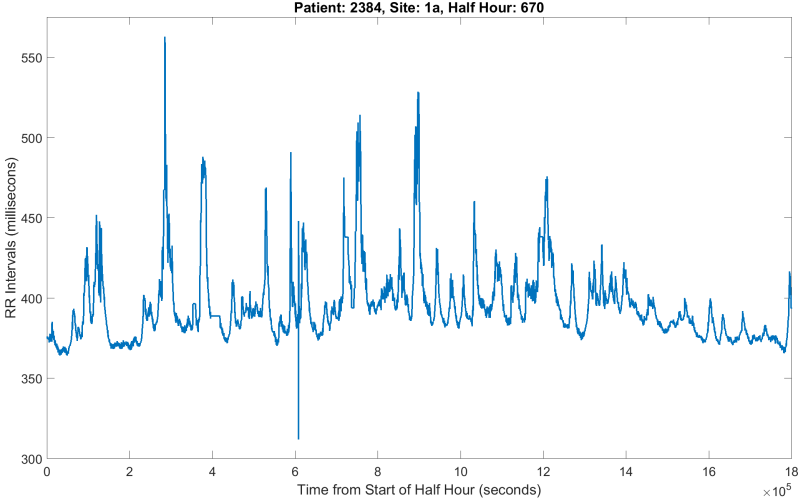
=



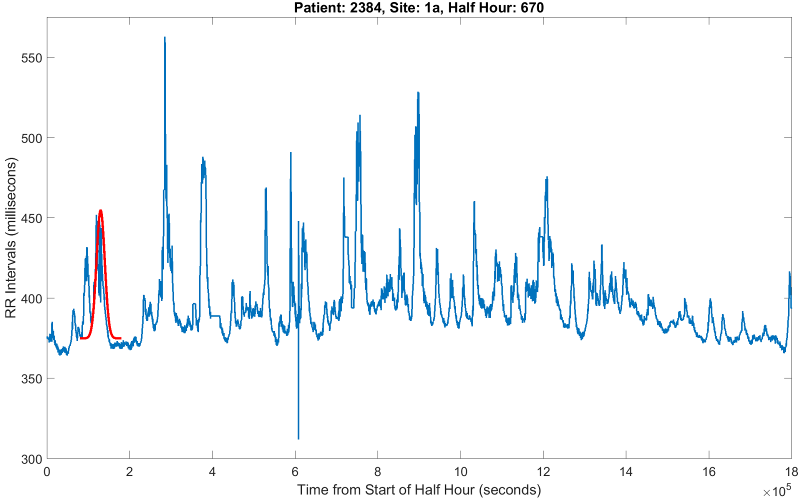
+



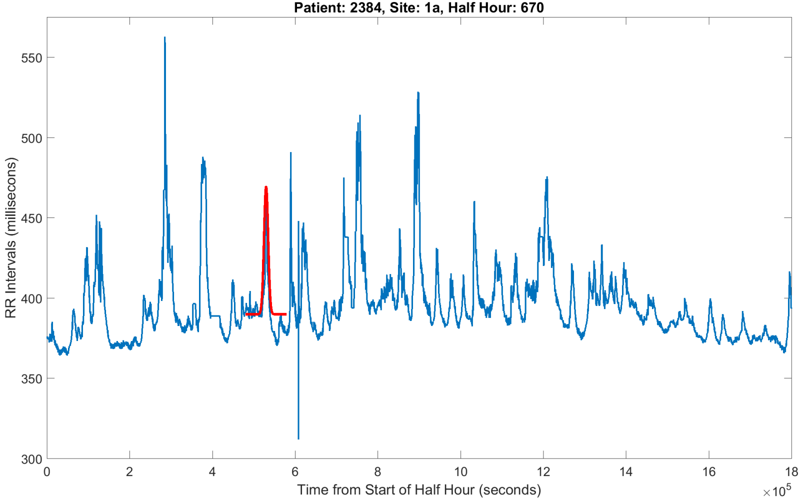
Heart Rate Characteristics : Decelerations



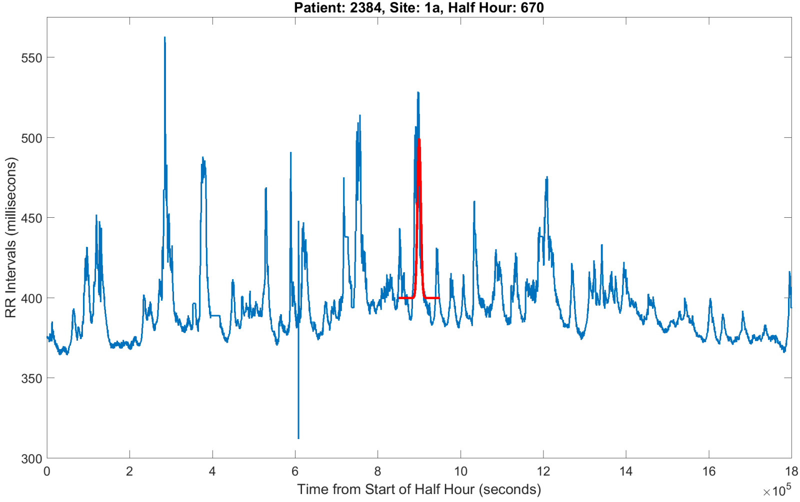
Heart Rate Characteristics : Decelerations



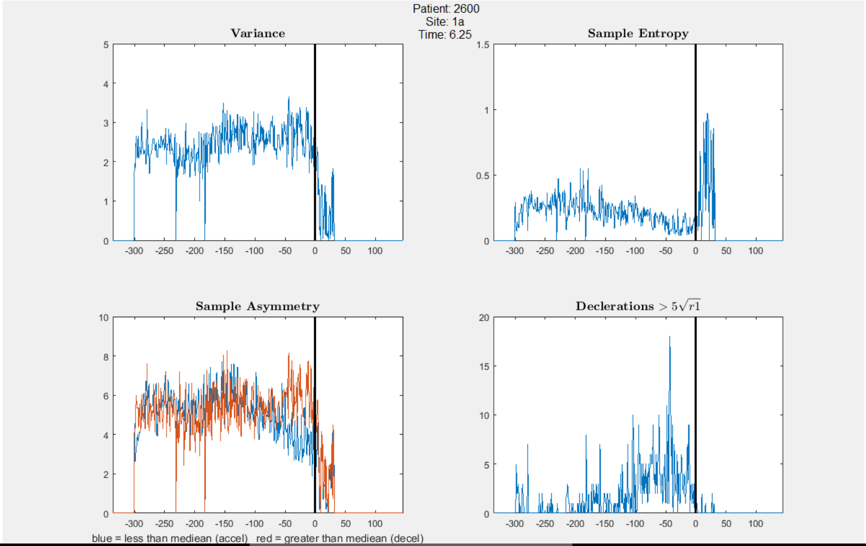
Heart Rate Characteristics : Decelerations



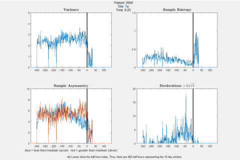
Heart Rate Characteristics : Decelerations



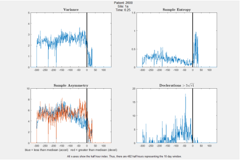
Heart Rate Characteristics

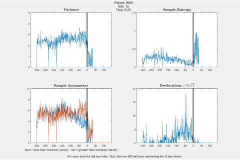


Heart Rate Characteristics

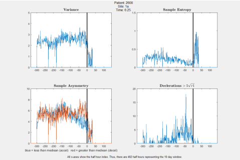


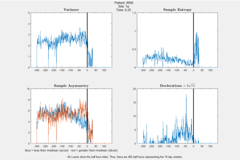
Heart Rate Characteristics

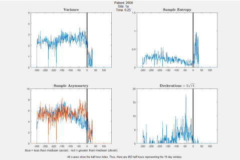




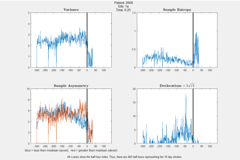
Heart Rate Characteristics

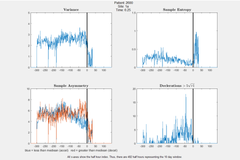


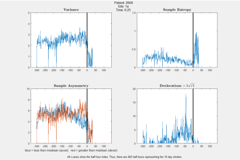


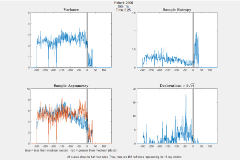


Heart Rate Characteristics

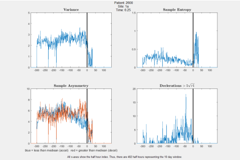


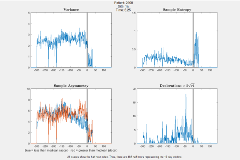


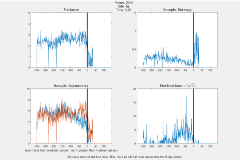


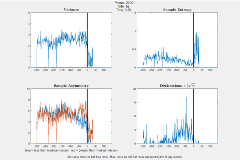


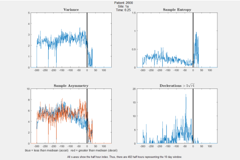
Heart Rate Characteristics



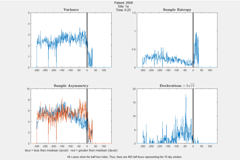


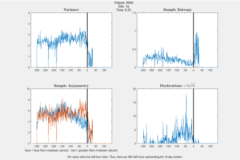


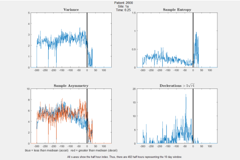


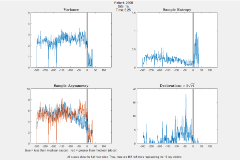


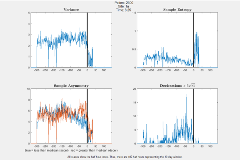
Heart Rate Characteristics

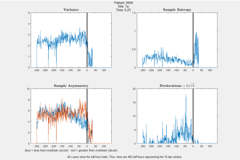




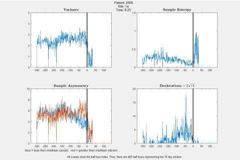


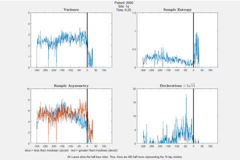


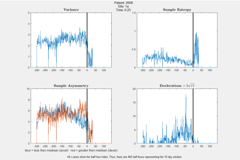


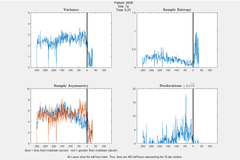


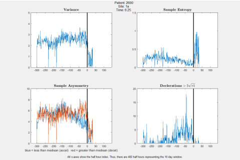
Heart Rate Characteristics

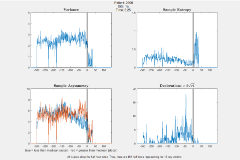


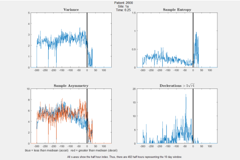




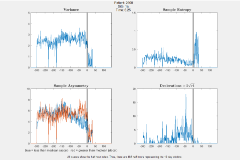
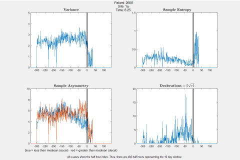
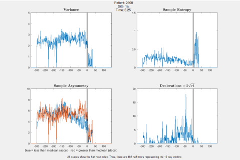
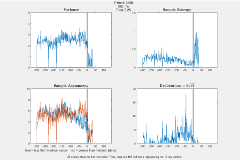
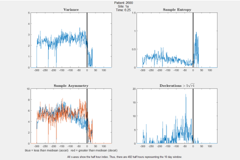
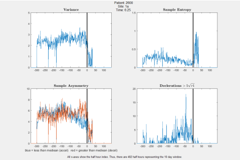
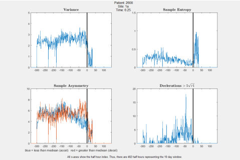
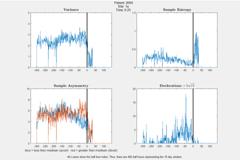




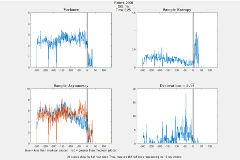
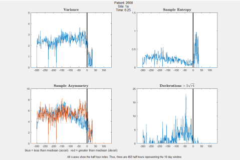
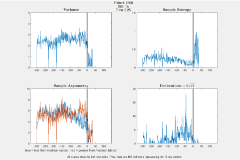
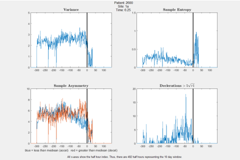
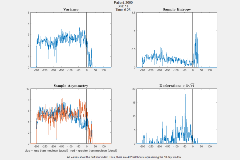
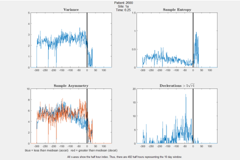
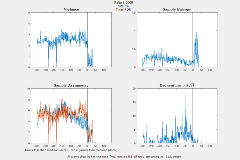
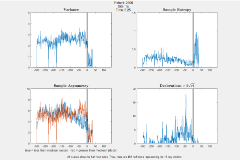




Heart Rate Characteristics

Heart Rate Characteristics

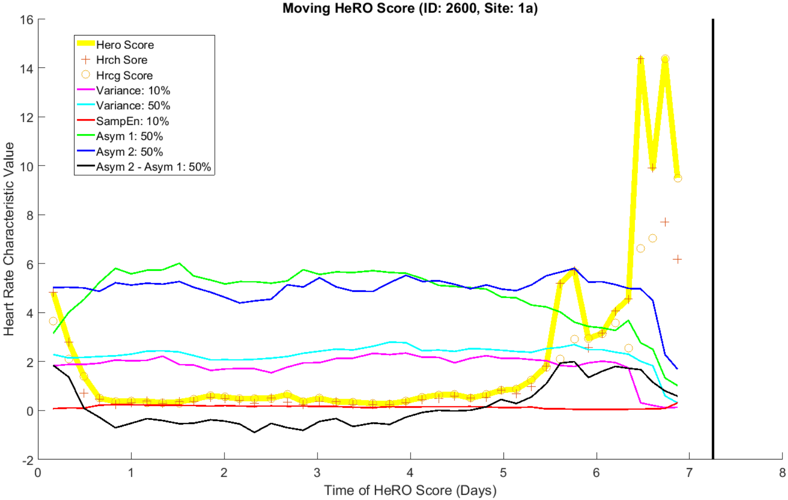
...

Analyzing the HRC's

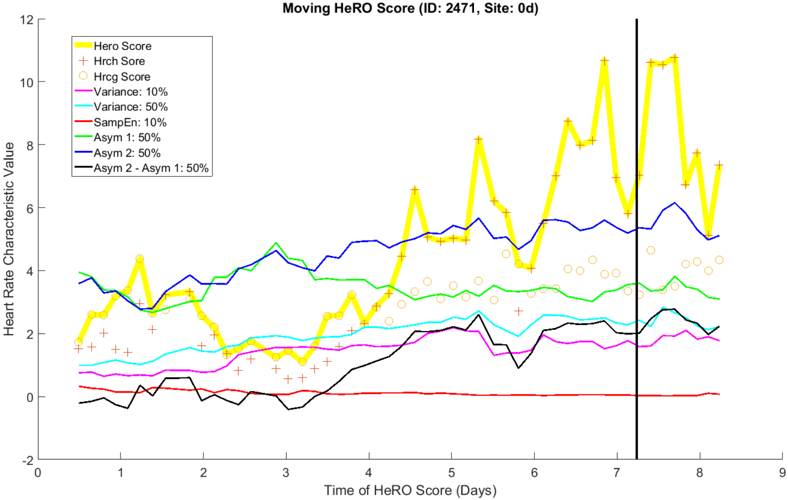
- HeRO Score
- Logistic Regression
- KNN
- Support Vector Machine

- HeRo Score
- Logistic Regression
- KNN
- Support Vector Machine

Analyzing the HRC's : HeRO Score



Analyzing the HRC's : HeRO Score



Analyzing the HRC's : Logistic Regression

- HeRo Score
- Logistic Regression
- KNN
- Support Vector Machine

Analyzing the HRC's : Logistic Regression, L1, L2, ELastic Net

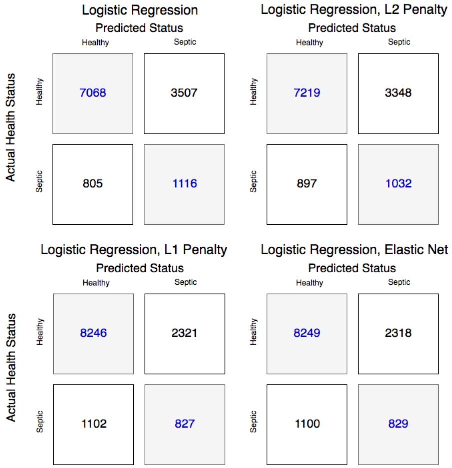


FIGURE – Logistic Regression with Penalties Confusion Matrices

Analyzing the HRC's : Logistic Regression, L1, L2, ELastic Net

Regression Method	Classification Accuracy	False Positive Rate
Logistic Regression	65.00%	64.61%
Logistic Regression, L1 Penalty	66.03%	63.45%
Logistic Regression, L2 Penalty	72.61%	54.61%
Logistic Regression, Elastic Net	72.65%	54.58%

TABLE – Summary of Results for Logistic Regression with Regularization

Analyzing the HRC's

- HeRo Score
- Logistic Regression
- KNN
- Support Vector Machine

Analyzing the HRC's : KNN Confusion Matrix

Actual Health Status \ Predicted Status	Healthy	Septic
Healthy	10349	261
Septic	1436	450

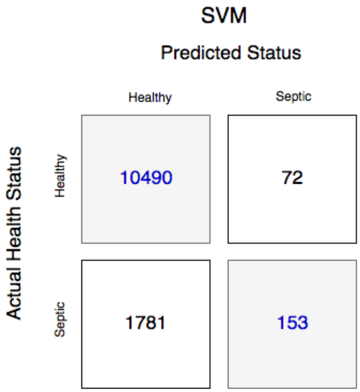
TABLE – KNN Confusion Matrix, 5-Fold Cross-Validation, Choose n-neighbors = 7

- Best performing predictors : log(variance), vent, sample entropy, decelerations
- Minkowski Distance, p=3 : $d(x^{(i)}, x^{j}) = \sqrt[3]{\sum_{k=1}^{n} | x_k^{(i)} - x_k^{(j)} |^3}$
- Classification Accuracy : $\frac{(10349+450)}{(10349+450+261+1436)} = 86.42\%$
- False Positive Rate : $\frac{261}{(261+450+1436)} = 12.16\%$

Analyzing the HRC's : Support Vector Machine

- HeRo Score
- Logistic Regression
- KNN
- Support Vector Machine

Analyzing the HRC's : Support Vector Machine



- Classification Accuracy : $\frac{(10490+153)}{(10490+72+1781+153)} = 85.17\%$
- False Positive Rate : $\frac{72}{(72+153+1781)} \approx 3.59\%$

Summary of Results

Regression Method	Classification Accuracy	False Positive Rate
Logistic Regression	65.00%	64.61%
Logistic Regression, L1 Penalty	66.03%	63.45%
Logistic Regression, L2 Penalty	72.61%	54.61%
Logistic Regression, Elastic Net	72.65%	54.58%
KNN, n=7	86.42%	12.16%
SVM, c=0.01	85.17%	3.59%

TABLE – Summary of Results

Organism Prediction

KNN, ORGANISM PREDICTION

Predicted Organism

Actual Organism	CONS	GRAM-POSTIVE	GRAM-NEGATIVE	FUNGAL	OTHER
CONS	791	47	60	38	2
GRAM-POSITIVE	47	349	30	8	3
GRAM-NEGATIVE	63	23	401	10	2
FUNGAL	43	12	11	145	0
OTHER	0	5	3	0	13

FIGURE – KNN, n=7 Organism Prediction Confusion Matrix

Summary

- Features : Variance, Sample Entropy, Sample Asymmetry, Decelerations, Ventilation Status, Weight, Gestional Age
- Methods : HeRO Score, Logistic Regression with L1, L2, Elastic Net Penalties, KNN, Support Vector Machine
- Results : We have promising results when determining if a baby is sick or healthy. However, if a baby is sick, determining the invading organism has proven to be more challenging.

Thank you for listening.

Thank you for listening.

Any questions ?

Thank you for listening.

Any questions ?

Acknowledgements : A special thanks to Professor John Delos, William and Mary Physics Department, and the University of Virginia NICU team for collecting the dataset and making this research possible.