Introduction

Sepsis—bacterial infection of the bloodstream—is one of the leading causes of death for preterm infants in neonatal intensive care units (NICUs) (source?). This is largely because sepsis can progress very quickly, yet symptoms may not be present until it is too late to administer antibiotics. The gold standard in the diagnosis of sepsis is a positive blood culture, but it can take several hours for a culture to come back positive, and again, it may be too late by that point.

This situation has incentivized researchers to develop tools for the early diagnosis of sepsis using vital sign measurements. One particular set of vital signs that have attracted interest are metrics of **heart rate variability [I think I need to give a better overview of this]**, meaning the variation in the amount of time between heart beats. Loosely speaking, heart rate variability is related to the nervous system's regulation of heart rate, which is in turn related to various immune response. The hope has been that heart rate variability shows changes in response to bacteria in the blood well before the bacteria proliferate to the point of inducing sepsis.

One well known tool in this field is the **HeRO score**, a proprietary algorithm which incorporates three heart rate variability metrics to produce a 1-to-5 score [or is it 0-to-5?] for the risk of developing sepsis [could use more details: developsing sepsis in the next how much time?] (Fairchild and Aschner). However, the HeRO score is not perfect, which has led other research groups to continue trying to solve the problem (ML article).

The purpose of this study was to calculate metrics on heart rate variability—mean RR, SDRR, RMSSD, pNN20, and pNN50, all to be defined—for a small set of infants, some septic and some not, and to answer this question: are there any clearly visible motifs in the time series of these signals that differentiate the septic infants from the nonseptic infants?

Materials and Methods

Dataset Description

The dataset for this study consisted of vital sign recordings from 7 preterm neonates, termed infants for simplicity.

Regarding selection criteria: The infants who developed sepsis, dubbed septic infants, were chosen based on the criteria that their presentations of sepsis were descried by a domain expert, Dr. Groves, as "very severe" and that they had no other issues within the recording period. The non-septic infants were chosen based on the criteria that they were close in age to one of the septic infants (so 1 goes with 6, 2 and 4 go with 5, and 3 goes with 2), and that they had "relatively uncomplicated" stays in the NICU.

Table 1 Overview of Infant Recordings

Infant	Became Septic	Gestational Age at Birth	Age at Start of Recording
1	Yes	27w 4d	29w 1d
2	No	26w 0d	27w 4d
3	No	24w 5d	33w 4d
4	No	25w 1d	34w 0d
5	Yes	24w 6d	34w 5d
6	No	28w 0d	29w 4d
7	Yes	24w 0d	27w 5d

In Table 1, gestational age at birth states the time passed from conception. Note that in a normal birth, the infant's gestational age is between 38 and 42 weeks.

The column sepsis time/center time is the time at the center of the recording. For the infants who became septic, the data was pulled such that the time at which a positive blood culture was recorded in the medical record was in the center. The center time is also calculated for the non-septic infants because this value is used to align the time series in the results section.

The vital sign data itself consisted of csv files whose columns were Time, at least one of ECG1, ECG2, and ECG3, and then Respiratory Impedence and SPO2. Because we were only interested in HRV, Respiratory Impedence and SPO2 were dropped. The ECG signals were sampled at 250 Hz. The reason we had up to three ECG signals is that three electrodes are required to record an ECG, and the bedside monitor recorded one of the signals at any given time. If the monitor lost the signal from one of the electrodes at any time, it could switch to another electrode. We found that the ECG signals ranged between similar values, and it was suggested by Dr. Chang that there was not a significant delay between the signals, so we decided to splice the signals together to create one signal, which we just called ECG.

The original files for each infant were enormous, so Dr. Chang broke each one into 5 files of manageable size (~3.5 GB), leading to a total of 35 files. The files were indexed by infant number and then "a" through "e", so, for example, "2b" would be a valid file index. This operation introduced an error for infants 2 and 3, where the end of the second file and the beginning of the third file was a jump in the time.

Figure 1 Gaps in the Time Indices for Infants 2 and 3

```
Last time in 2b 863999.9998149872
First time in 2c 5270400.157104015
Last time in 3b 5082017.614679575
First time in 3c 5271941.108734131
```

There were segments of missing data in all of the ECG signals, sometimes placed sporadically, but often due to the monitor switching between ECG signals. There were also sporadic spikes in all of the signals—likely due to infant movement—and strange artefacts called troughs which needed to be addressed.

To summarize, there are some issue with the reliability of this data, such as the time index gap mentioned above, as well as an issue of the statistics for infants 2 and 4 that is best illustrated in the results section. This data was originally collected for the purpose of monitoring the infants' health in the NICU. This dataset is not sufficient for the purpose of making inferences because it is too small and because the criteria for inclusion and exclusion are not defined precisely enough, although that is acceptable in this case because the purpose of this project was simply to produce visualizations. This data is current, with the oldest infant in the dataset born after 2020. Finally, this data is housed by Dell Medical School at the University of Texas at Austin.

Regarding licensing, I completed IRB training to ensure that I had the necessary credentials to access this data. For privacy purposes, the data has been deidentified and it does not disclose any personal information other than gestational age at birth. As stated above, this data is securely housed by Dell Medical School, so it can only be accessed by someone affiliated with that institution. My access to this data was through UT Box.

At the outset, we believed that this dataset had the potential to answer our question by showing visual differences between HRV statistics for septic and non-septic infants, if such differences exist. However, the inconclusive result of this study indicates that a more careful selection of data is necessary.

Heart Rate Variability Metrics

There are many heart rate variability metrics. At a broad level, they can be classified as time-domain, frequency-domain, and nonlinear (source about overview). In this study, we chose to only calculate a selection of time-domain metrics. This decision was based in part on a finding that time-domain metrics vary significantly between septic and non-septic infants, whereas frequency domain metrics do not (source). The other factor in the decision was the principle of Occam's razor; time-domain metrics are simpler to calculate than frequency-domain and nonlinear metrics, and since this was our first investigation into this field, we wanted to investigate simpler metrics first.

The metrics suggested by Dr. Groves were:

- The mean RR interval, or meanRR
- The standard deviation of the RR intervals, or SDRR
- The root mean square of successive differences, or RMSSD
- The probability of two RR intervals in a row differing by more than 50 ms, or pNN50
- The probability of two RR intervals in a row differing by more than 20 ms, or pNN20

Note that meanRR is technically not an HRV metric, but it was so simple to calculate that we included it in case it revealed anything interesting.

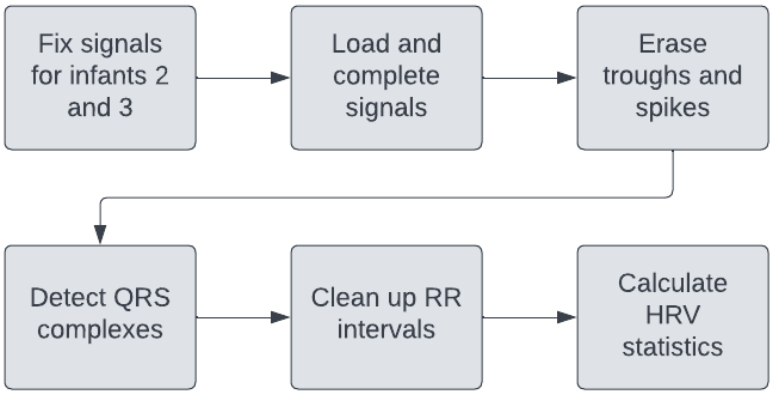
The statistics were calculated using a rolling window. We used window lengths of a week, a day, and an hour, and found that a day struck a good balance between being too noisy and too flat.

QRS Detector

The term **QRS complex** refers to the characteristic heart beat curve one sees in ECG data, and the location in time of the peak of the QRS complex is called the **R peak**. In order to calculate heart rate variability statistics, one needs to know the location in time of every R peak in the signal. Labeling R peaks by hand with this amount of data would be infeasibile, so we needed an algorithm called a **QRS detector** to do this labeling for us. Of the many QRS detectors available, we chose to use **XQRS** in the **Waveform Database (WFDB)** Python library. There are two reasons for this. First, a research paper showed that XQRS was one of the top three QRS detectors for noiseless ECG signals from adults (source). Although we are working with noisy signals from infants, this seemed like a good place to start. Second, it was one of the two QRS detectors readily available, and the other one, GQRS, is no longer updated.

Data Pipeline

Figure 2 Data Pipeline



Step 1: Fix signals for infants 2 and 3

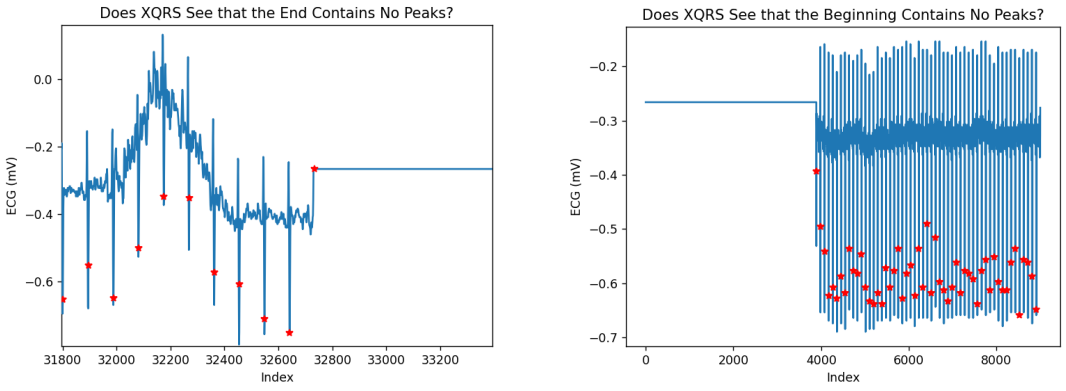
To resolve the time index gap problem described in the Dataset Description section, we shifted the time indices for files 2a and 2b to line up with 2c, and then 3a and 3b to line up with 3c. Because the data was deidentified, the time index does not correspond to the birth date of the infant, so performing such a shift preserves data integrity.

Step 2: Load and complete signals

Due to the size of the files, each file was loaded one at a time. The respiratory impedance and SPO2 data was dropped. Then, we constructed one ECG signal out of the three channels in the following way: First, we took ECG1. If there was data for ECG2, then we used that to fill in any remaining missing values. If there was data for ECG3, then that data was used to fill in any remaining missing values. If there were any more missing values after the signals were combined, then those were filled in using fill forward, meaning that the last non-missing value was written over the missing values until a new non-missing value was reached.

Using forward fill was necessary because XQRS would fail to detect any R peaks after a segment of missing data. Moreover, XQRS would not detect any R peaks in a filled in region, so ine effect, the data was still interpreted by XQRS as missing. See Figure 3 for an illustration.

Figure 3 How XQRS Responds to a Segment of No R Peaks



In this figure, observe that XQRS is able to identify QRS complexes both before and after a segment of missing data.

Step 3: Erase troughs and spikes

Figure 4 Example of Spikes and Troughs

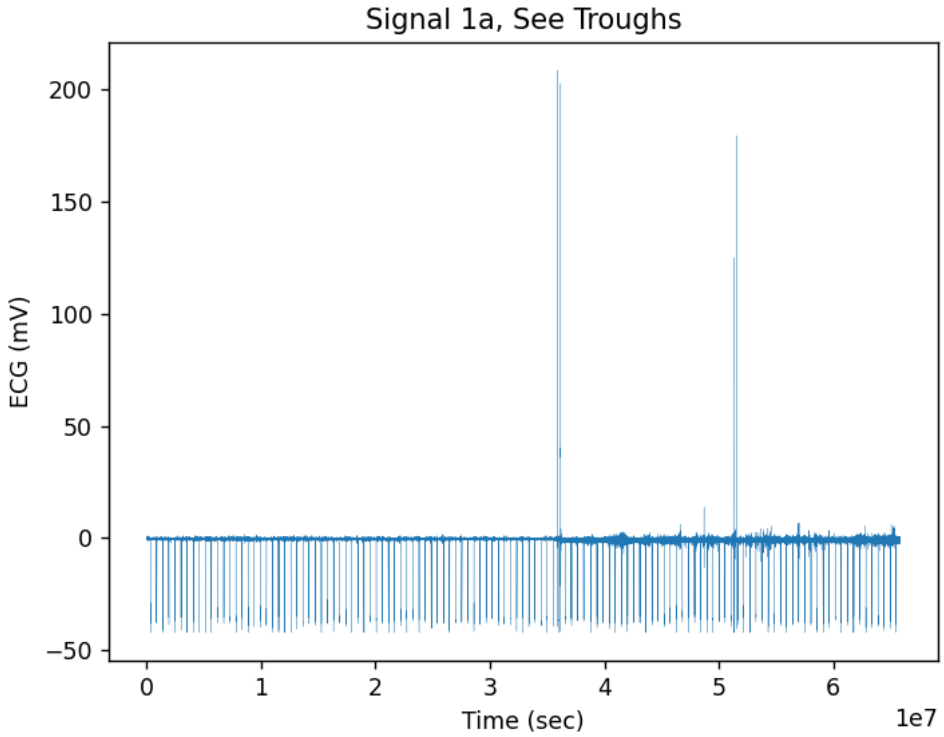


Figure 4 shows the ECG signal constructed from file 1a. The majority of the signal is near 0 mV, which is an appropriate range for the ECG signal. However, the signal also shows two kinds of major deviations from the baseline, spikes and troughs. By spikes, we are referring to the large positive jumps in the signal, and by trough, we are referring to the somewhat periodic dips below baseline that reach to approximately -50 mV. Spikes typically shoot straight up and down in a simple fashion, but troughs are more complex. Figure 5 shows an example of the first trough in Figure 4. The x-asix has been shrunk so that the fine structure of the trough is visible, ant the y-axis has been set to contain the baseline. Observe two things: first, the signal rapidly oscillates between baseline and a very negative value, and second, this oscillation obscures QRS complexes. This second point has two negative effects: one, it artificially increases the RR interval, and two, XQRS fails to identify R peaks following a trough. Because of this, we decided to erase any part of the signal corresponding to a trough, and then fill it in will fill forward. Due to the artificial spacing out of the QRS complexes, we erase the entire trough, not just the parts that deviated significantly from baseline. Figure 6 shows the signal before and after trough erasure.

Figure 5 Zoom of a trough

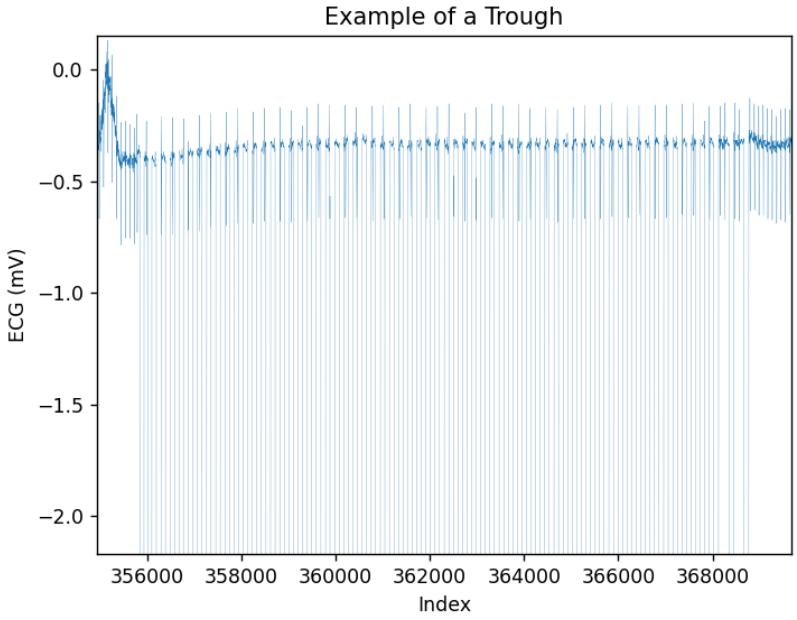
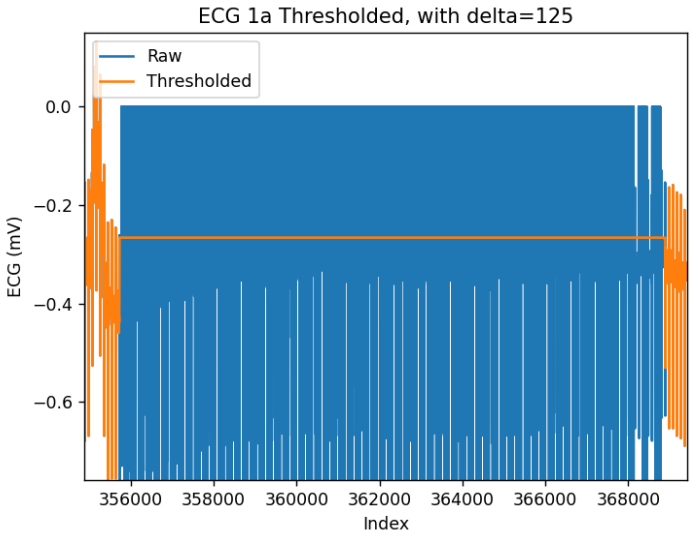


Figure 6 Signal before and after trough erasure



Step 4: Detect QRS Complexes

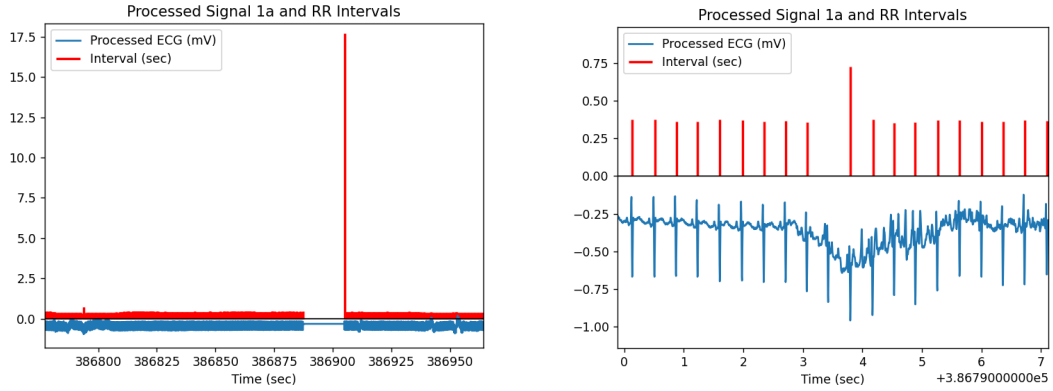
Once troughs are removed, the signal is clean enough to be passed to XQRS. We found that applying XQRS to the whole signal all at once had an unacceptably slow runtime, but that it performed more quickly on smaller segments, so we decided to break each signal into chunks of 10,000 indices, detect QRS complexes in each chunk, and then combine all of the R peak locations into one output file.

Step 5: Clean up the RR intervals

Once the R peaks have been detected, the RR intervals can be calculated by taking the difference between each R peak and the last. Because the first R peak has no previous R peak, we do lose one datapoint, but this is an insignificant amount of data loss.

With the RR intervals in hand, there's still a little more data cleaning to be done because some RR intervals are too short or too long. Dr. Groves suggested that the physiological lower bound for an RR interval is 0.25 seconds, so any intervals after that were removed. For the RR intervals that were too long, there were two causes: missing data and undetected beats. We termed lengthy RR intervals due to missing data **missing intervals** and lengthy RR intervals due to undetected beats **multiple intervals**.

Figure 7 Examples of RR intervals that are too long



The left panel of Figure 7 shows a missing interval, while the figure on the right shows a multiple interval; in this figure, it appears that missing intervals can be distinguished from multiple intervals by their larger length, but in practice, it was not so simple. By visual inspection, we found that XQRS could miss up to 17 beats in a row, which would lead to a very long multiple interval.

We wanted to recover some of the beats lost to multiple intervals, so we attempted to do so as follows. First, we arbitrarily decided that any interval longer than 5 seconds was a missing interval, and threw it away. For any interval shorter than 5 seconds, we approximated how many time longer it was than the previous interval, and then broke it into that many pieces to recover single intervals. This method is suboptimal, as is discussed in the discussion section.

Step 6: Calculate HRV Metrics

After step 5, we decided that the RR intervals were clean enough to calculate HRV metrics, which are in the results section.

Results

Unfortunately, all of the statistics were inconclusive. No visual indicators of sepsis were visible. Still, we give the figures showing this.

- Aligned at 0
- Note that the signal for infant 2 vanishes behind that of infant 4 in every plot. This points to a programming error.
- Note that RMSSD, pNN20, and pNN50 are all correlated.

Figure 8

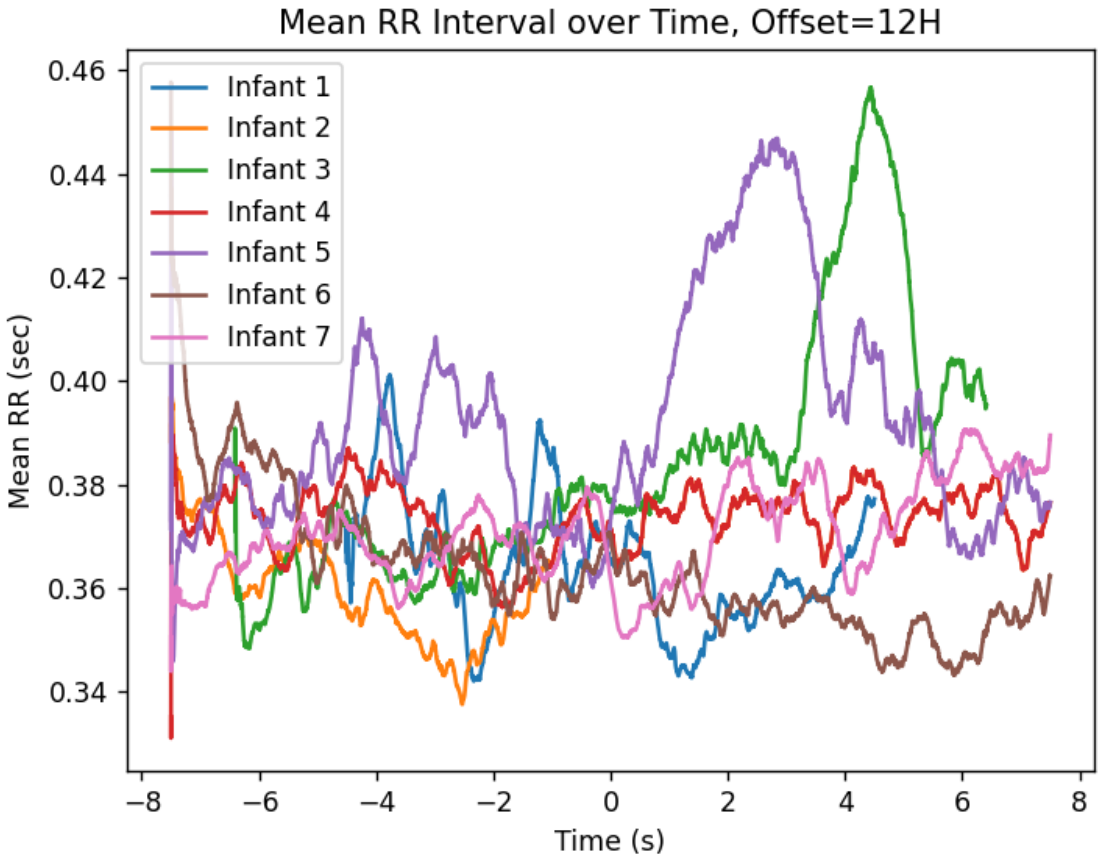


Figure 9

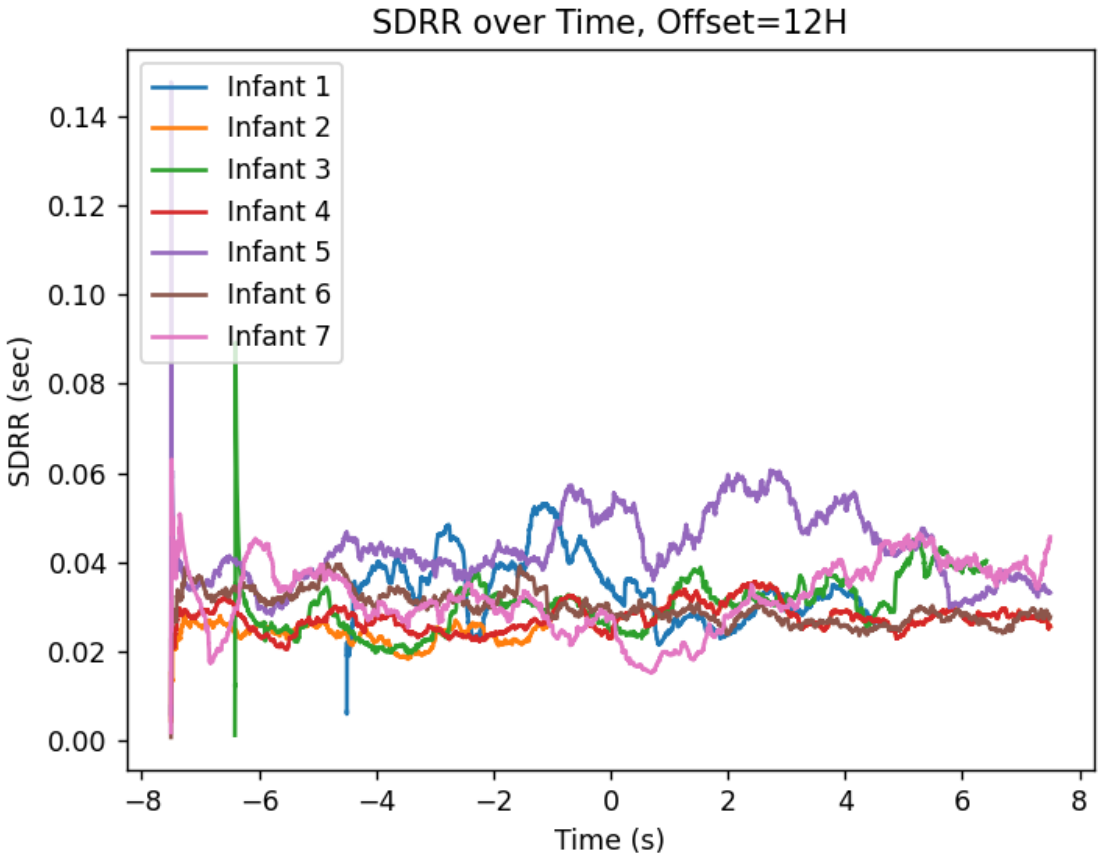

SDRR over Time, Offset=12H
Infant 1
Infant 2
Infant 3
Infant 4
Infant 5
Infant 6
Infant 7
SDRR (sec)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−8
−6
−4
−2
0
2
4
6
8
Time (s)

Figure 10

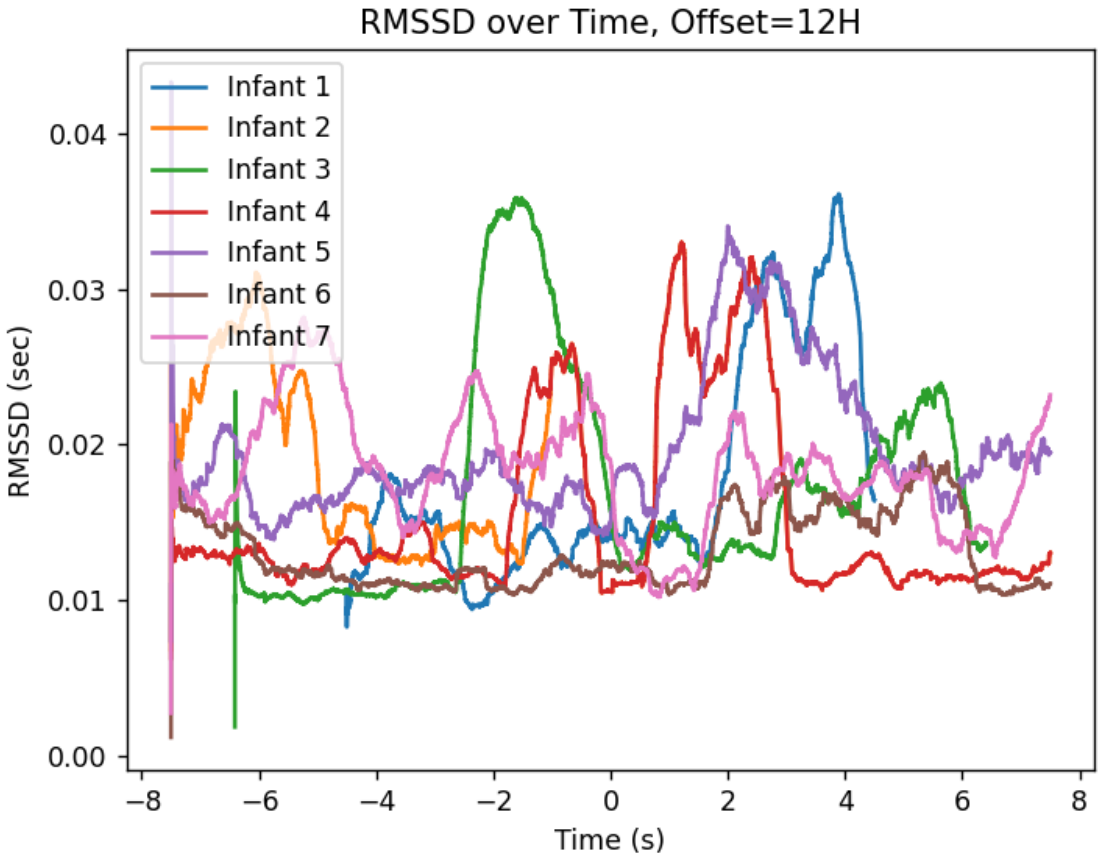

RMSSD over Time, Offset=12H
Infant 1
Infant 2
Infant 3
Infant 4
Infant 5
Infant 6
Infant 7
RMSSD (sec)
0.04
0.03
0.02
0.01
0.00
−8
−6
−4
−2
0
2
4
6
8
Time (s)

Figure 11

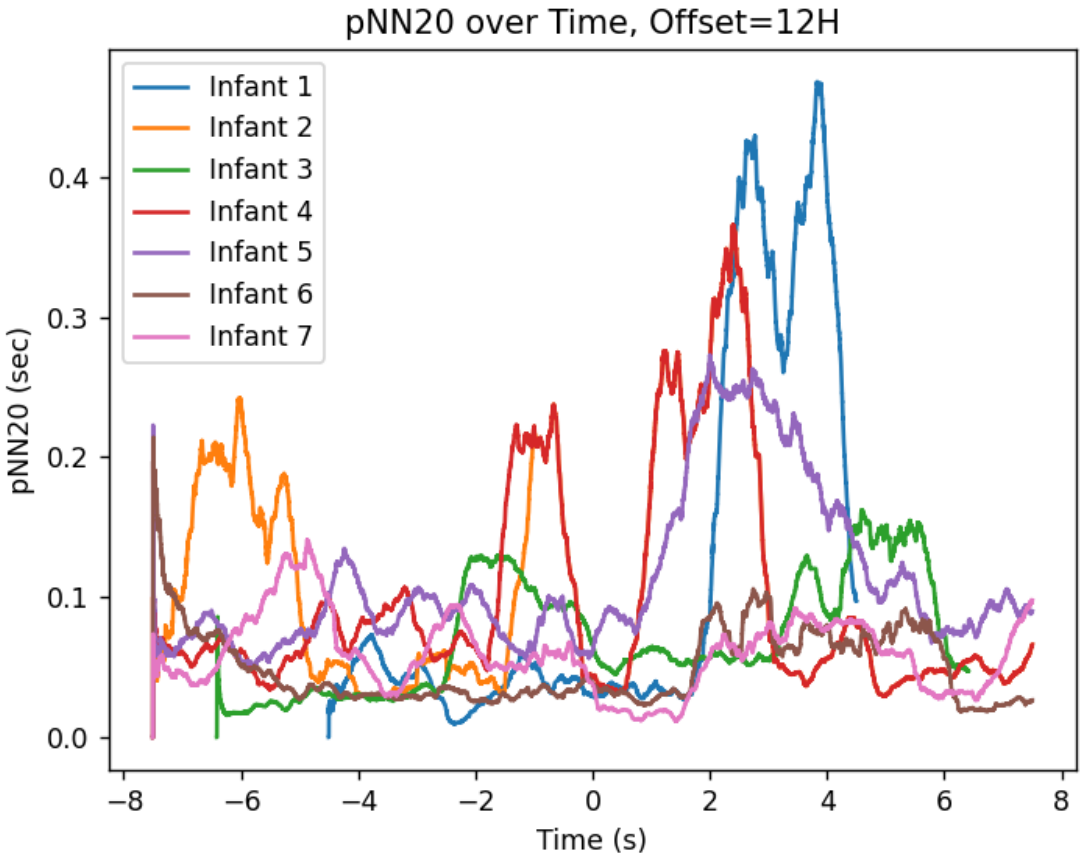
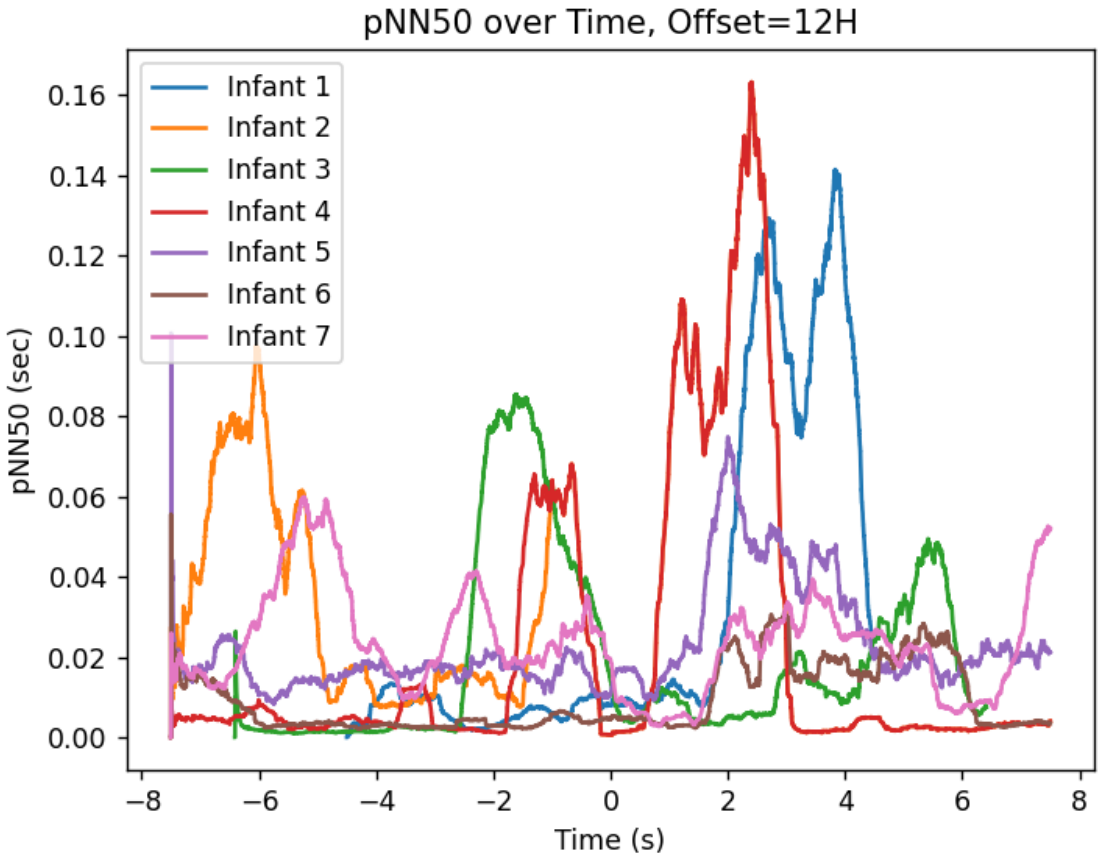

pNN20 over Time, Offset=12H
Infant 1
Infant 2
Infant 3
Infant 4
Infant 5
Infant 6
Infant 7
pNN20 (sec)
0.0
0.1
0.2
0.3
0.4
−8
−6
−4
−2
0
2
4
6
8
Time (s)

Figure 12



Discussion

- We need a more careful pulling of the data and file management to make sure that the 2-4 overlap problem is fixed
- A more careful query would also help with the time index gap for infants 2 and 3.
- We tried to ascertain the cause of the troughs, but nothing came to mind. It seems to be a problem with the recording device, which would require someone in the NICU to inspect.
- Applying a more sophisticated denoising to the signal that decreases the baseline whitenoise may improve QRS detection and reduce the need for breaking up multiple intervals further down the pipeline
- The method of breaking up RR intervals assigns equal length to each interval, which artificially deflates the variability. Because the rolling window was over 12 hours, we found this to be acceptable, but because the results was inconclusive, other rolling windows may be tried, hence a more sophisticated method for breaking multiple intervals that preserves some of the variability may be helpful.

Acknowledgements

I would like to thank Dr. Joshua Chang and Dr. Alan Groves of the University of Texas at Austin for their support throughout this projects, including providing data, programming tips, and troubleshooting suggestions, and listening to presentations along the way.

This work was done for the course UGS 320K: Undergrad Research Experience in the Fall of 2022 at the University of Texas at Austin.

Bibliography

Bohanon, F. J. et al, (2015). Heart rate variability analysis is more sensitive at identifying neonatal sepsis than conventional vital signs. The American Journal of Surgery, 210(4), 661–667. https://doi.org/10.1016/j.amjsurg.2015.06.002

Eilers, J., Chromik, J. & Arnrich, B. (2021). Choosing the appropriate QRS detector. Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, 4, 50–59. https://doi.org/10.5220/0010234600500059

Fairchild, K. D. & Aschner, J. L. (2012). Hero monitoring to reduce mortality in NICU patients. Research and Reports in Neonatology, 65–76. https://doi.org/10.2147/rrn.s32570

Masino, A. J. et al (2019). Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. PLOS ONE, 14(2). https://doi.org/10.1371/journal.pone.0212665

Shaffer, F. & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. Frontiers in Public Health, 5. https://doi.org/10.3389/fpubh.2017.00258