Blood Pressure Analysis with PPG and ABP

By: Franklin Zhu

Purpose

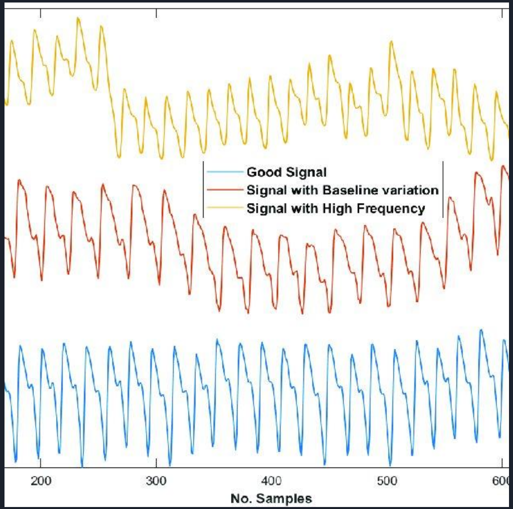
Issues:

- Blood pressure is a key indicator of several conditions and diseases
- Limitations of blood pressure (BP) measuring devices

Goal:

- Accurately predict blood pressure from PPG signals
- Easy to measure and use
- Various machine learning methods have already been implemented

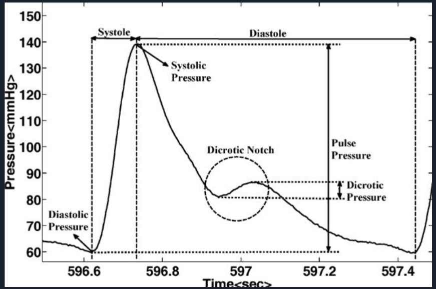
What are PPG and ABP signals?



Photoplethysmogram (PPG) signals are waves that consist of a systolic and diastolic peak and a foot.

- Measures volumetric changes in blood in circulation

Arterial blood pressure (ABP) signals have a structure very similar to PPG, with a systolic and diastolic peak.



MIMIC-III File Format

- Each patient directory consists of various .hea and .dat files
- Every file is for different segments of signals
- Layout .hea file specifies which file contains which signal
- Filtered for directories that contain both pleth and abp signals
- Used wfdb2mat to convert all .dat files into .mat

3544749_0007	12/31/1999 9:00 PM	HEA File	1 KB
3544749_0008	12/31/1999 9:00 PM	DAT File	5,303 KB
3544749_0008	12/31/1999 9:00 PM	HEA File	1 KB
3544749_layout	7/1/2017 5:00 AM	HEA File	1 KB
3544749n	12/31/1999 9:00 PM	DAT File	39 KB
index	6/13/2022 10:29 AM	Chrome HTML Docu...	3 KB
p000020-2183-04-28-17-47	7/1/2017 5:00 AM	HEA File	1 KB
p000020-2183-04-28-17-47n	7/1/2017 5:00 AM	HEA File	1 KB
RECORDS	8/3/2017 10:03 PM	File	1 KB

Data Processing

Steps (mainly follows same processing as Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network):

1. Convert .hea and .dat files to .mat with wfdb2mat
2. Filter for files that have both ABP and PPG waveforms
3. Ignore graphs that have flat waveforms
4. Apply Butterworth band-pass filter to eliminate noise below 0.5Hz and above 8Hz
5. Apply Hampel filter to remove outliers

Downloading Data

- Downloaded data from MIMIC-III waveform database matched subset
- Used a python os to go through all files and convert them to .mat in cygwin
- Removed files that did not contain necessary waveforms (specified as PLETH and ABP)

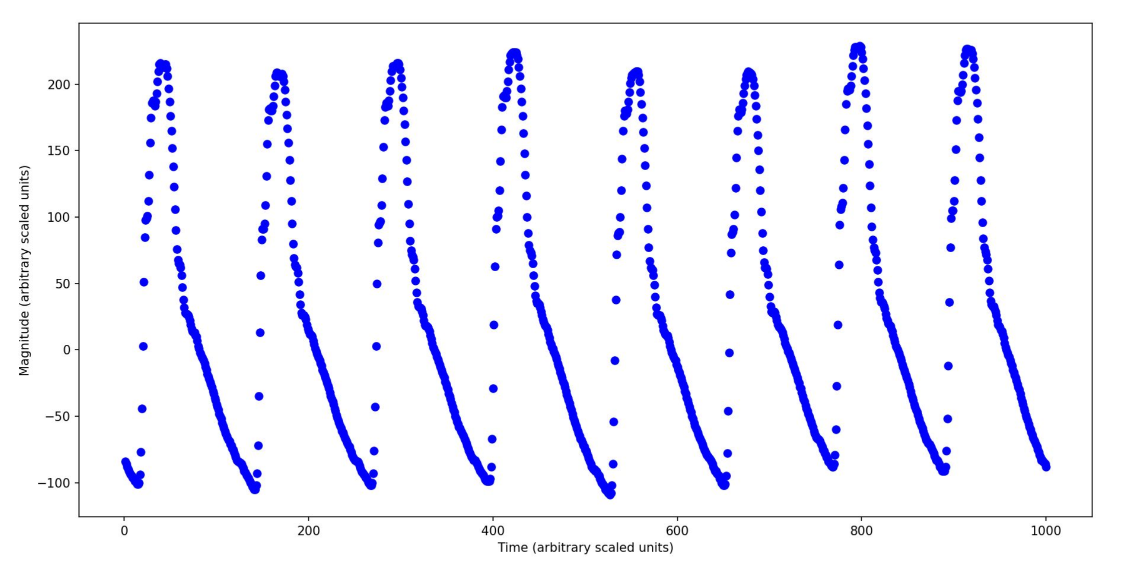
.hea file example

```
p087119-2156-01-04-17-25m 3 125 38797500 17:25:55 04/01/2156
p087119-2156-01-04-17-25m.mat 16+192 127(-64)/mV 8 0 -32768 -25071 0 II
p087119-2156-01-04-17-25m.mat 16+192 255(-128)/NU 8 0 -32768 -7335 0 PLETH
p087119-2156-01-04-17-25m.mat 16+192 4.825(-410)/mmHg 10 0 -32768 31134 0 ABP
# Location: tsicu
#Creator: wfdb2mat
#Source: record p087119-2156-01-04-17-25  Start: [17:25:55.000 04/01/2156]
```

.mat File Format

- Wfdb2mat converts .dat files into .mat files with a .hea descriptor file
- .mat file can be loaded with scipy.io.loadmat as an array of arrays
- Resulted in weird values

Ex: Plotted BP array has peaks over 200

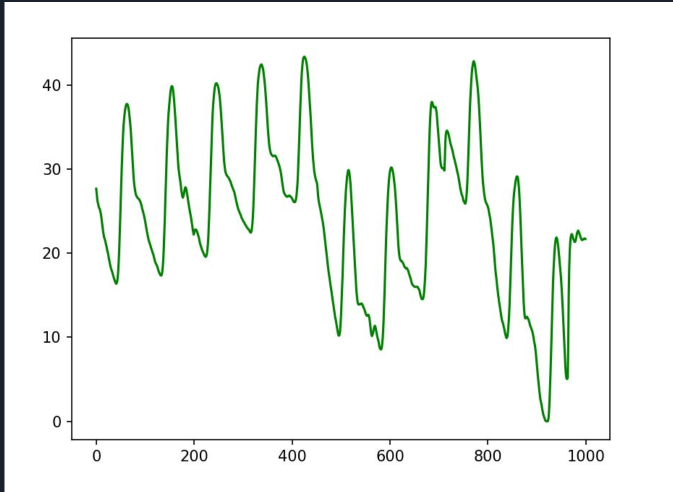


Solution

- Ended up reading .mat files with wfdb python package
- Wfdb.rdrecord creates a record object from .hea and .mat file
- Used p_signal property of record class to read an array with values for each signal
- This resulted in proper graphs
- Not sure why directly reading from .mat file produced different results

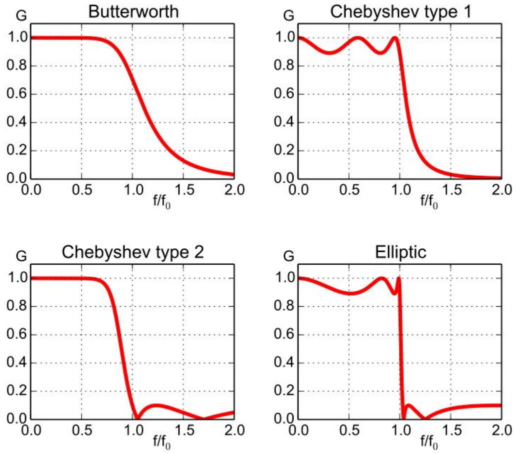
MIMIC-III Signal Problems

- Not all signals available in the same interval
- Sometimes PPG is present, but with no ABP and vice versa
- Some segments contain both signals, but one/both of them have problems
- Lots of PPG signals had high frequency noise or unnatural peaks
- Lots of flat lines, as signals were not available at all times
- Did not experience flat peaks, as described by previous paper



Filtering Techniques

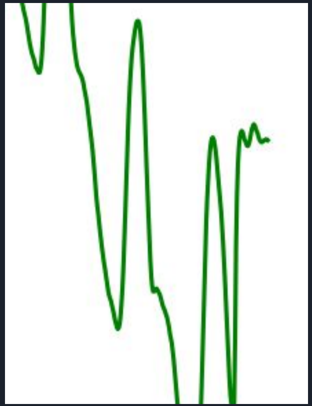
- Previous mentioned paper applied 4th order Butterworth bandpass filter with cutoff frequencies of 0.5Hz and 8Hz
- Tried to implement this with scipy.signal.butter
- Not completely sure why, but high frequency areas were not flattened out properly



The Butterworth filter rolls off more slowly around the cutoff frequency than the Chebyshev filter or the Elliptic filter, but without ripple.

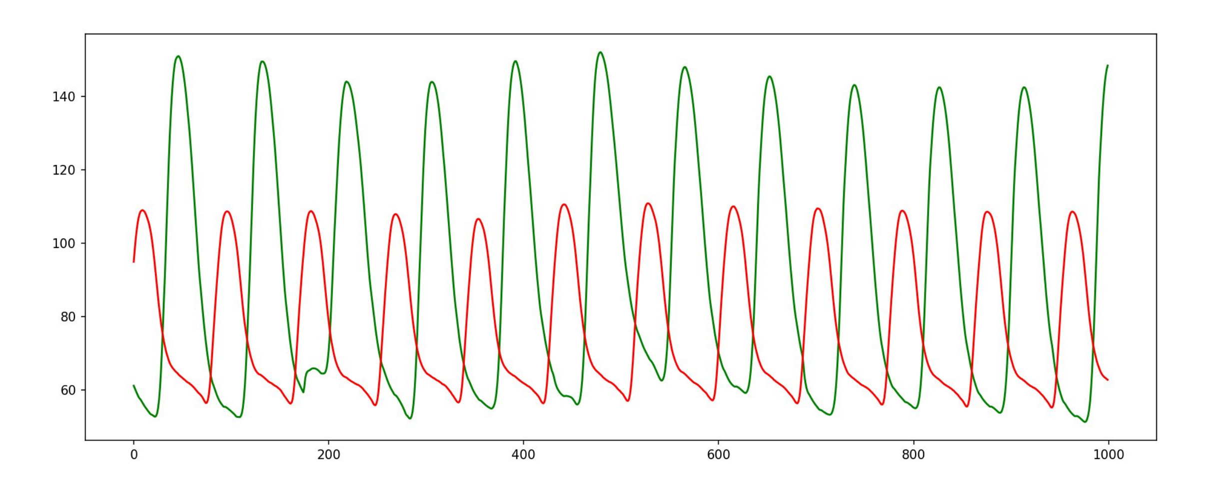
Filtering Techniques cont.

- Previous paper also applied Hampel filter
- Hampel filter uses a rolling median to remove outliers and replace them with more reasonable values
- Tried to implement this and tested with multiple window sizes, but the filter was not effective
- Main problem with PPG signal was the high frequency of small peaks, which were not outliers by magnitude



Solution

- Could not filter data in a way to remove bad sections
- Had to visually inspect segments of 10,000+ points from patient files
- Chose 8 patients with pretty good PPG and ABP waveforms



Red: ABP, Green: Scaled PPG

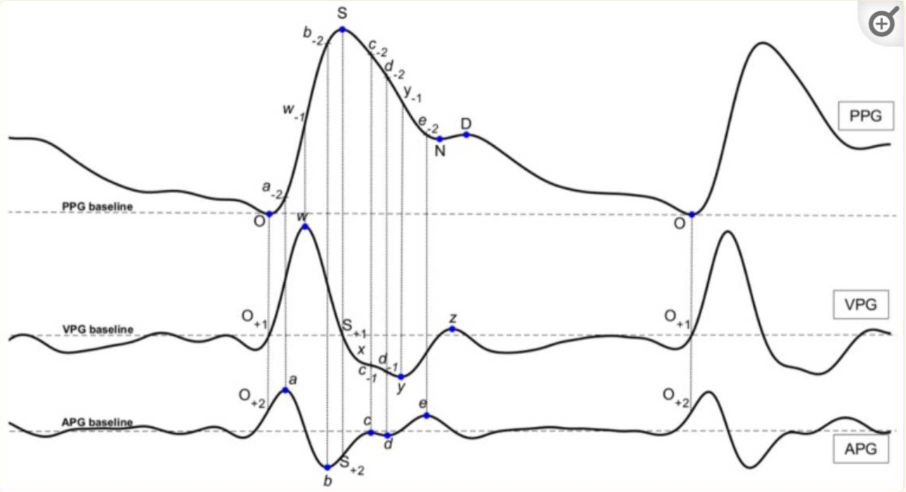
Features of PPG Waveforms

- Many components of PPG waves are correlated to BP
- Labeled features are shown in the image (taken from Hypertension Assessment Using Photoplethysmography: A Risk Stratification Approach)
- Top 5 correlated features from same study shown below

Table 1

A correlation coefficient list of top 10 features with systolic blood pressure (SBP).

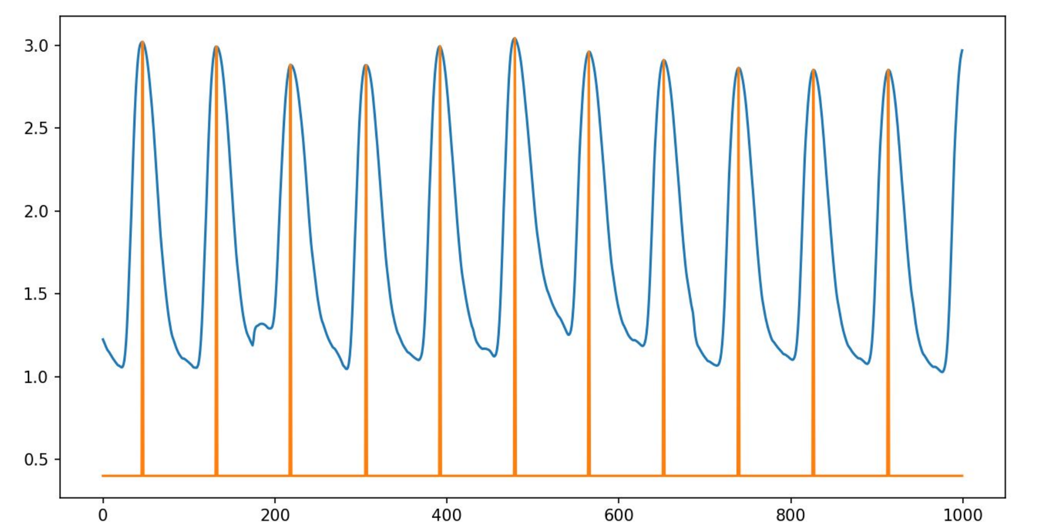
Index	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Feature	$(b\text{-}c\text{-}d)/a$	$\widetilde{b_{-2}d_{-2}}$	$\widetilde{\mathbf{S}c_{-2}}$	$\overline{\mathbf{OS}}$	$c_{-2}/\mathbf{S}$	$\overbrace{\mathbf{S}_{+1}\mathbf{c}_{-1}}$	$\overbrace{\mathbf{S}_{+1}\mathbf{c}_{-1}}/\overbrace{\mathbf{O}_{+1}\mathbf{O}_{+1}}$	$b_{-2}/\mathbf{S}$	$\overline{\mathbf{S}\mathbf{c}_{-2}}$	$\overbrace{\mathbf{wS}_{+1}}/\overbrace{\mathbf{O}_{+1}\mathbf{O}_{+1}}$
Correlation coefficient	0.6903	0.6721	0.6181	0.6164	0.5332	−0.4722	−0.5928	−0.6353	−0.6391	−0.6482
p-value	0.0001	0.0009	0.0013	0.0017	0.0425	0.0443	0.0058	0.0004	0.00003	0.0079



Calculating Crest Time

- Found when a given signal was increasing
- Marked points where it changed from increasing to decreasing
- Used a threshold to eliminate small peaks under it and used the frequency (125 Hz) to locate peaks for a single period and find the maximum

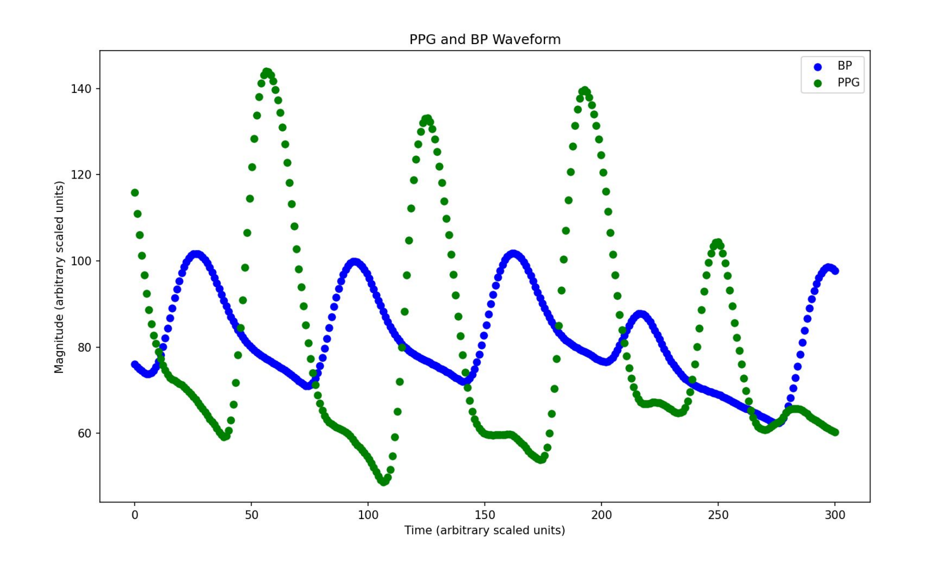
- Found minimum value between 2 peaks as the foot
- Calculated time between the foot and the peak to find CT



Calculating SBP

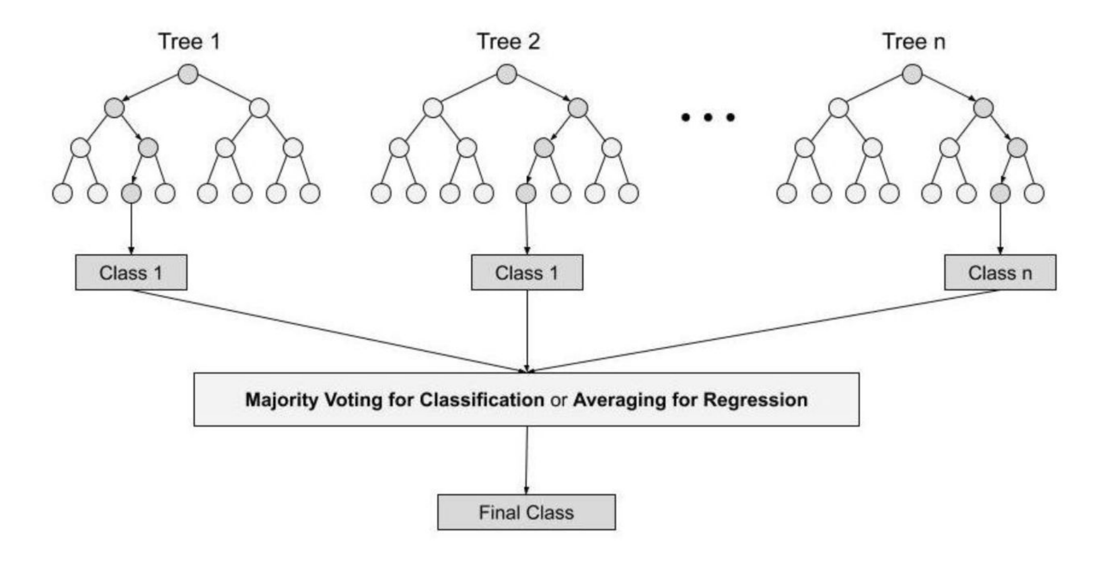
- SBP is the peak of ABP waveform
- Correspond each CT with BP

- Peaks alternate with PPG peaks
- Take time window between 2 PPG peaks
- Find the maximum value in the time window



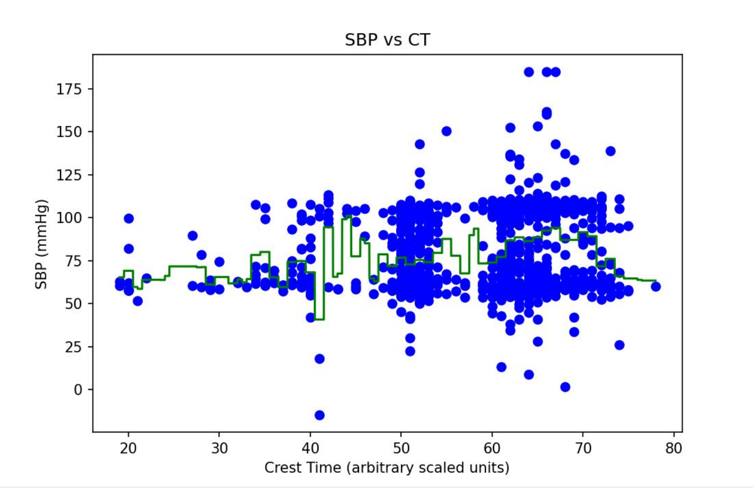
How does Random Forest work?

- Uses random decision trees picking random separators at every node
- Uses majority voting of decision trees to make final decision
- Used RandomForestRegressor from sklearn



Correlation

- Correlation coefficient obtained was unusually low compared to other paper (0.3525 vs 0.6164)
- Possibly due to problems in data filtering and smaller sample size
- Thus, random forest results were not great
- Did not have time to precisely configure random forest, but another possible issue with predictions could be overfitting



Conclusion

- Largest obstacle is processing data
- Lots of potential for PPG prediction
- Optimize random forest