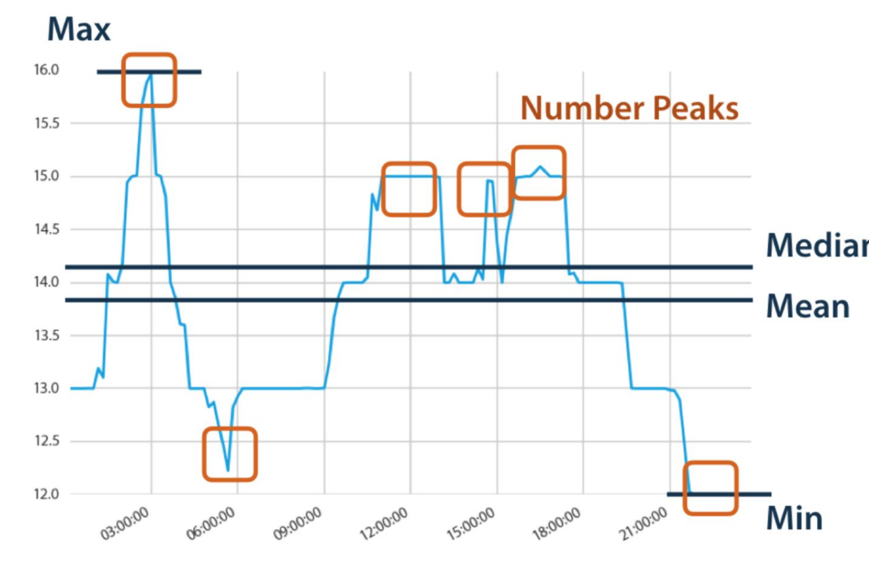
Feature Engineering with High-Frequency Physiologic Data

Akash Bhandari

Introduction:

Fever is defined as a state of having body temperature >= 38C. Diagnosis of fever is a major challenging task to the physicians which often remains undiagnosed and delays the treatment. Therefore, monitoring of fever provides valuable information for diagnosis and prognosis of the disease. Previously, Convolutional Neural Networks and different machine learning models have been used to predict dengue fever where surrounding atmospheric data (humidity, temperature, precipitation etc.) was included. In the study, features are extracted and a model is built to predict the onset of fever on ICU patients using the high frequency physiologic waveform data.



1

	ID	Date	MDC_PRESS_CUFF_SYS	MDC_PRESS_CUFF_MEAN	MDC_PRESS_CUFF_DIA	MDC_PULS_OXIM_SAT_O2	MDC_PULS_OXIM_PULS_RATE	MDC_EC
0	2330393	2018-01-29 09:16:03	85.0	73.0	61.0	NaN	NaN	
1	2330393	2018-01-29 12:26:31	NaN	NaN	NaN	99.0	129.0	
2	2330393	2018-01-29 12:27:31	NaN	NaN	NaN	100.0	130.0	
3	2330393	2018-01-29 12:28:31	NaN	NaN	NaN	99.0	126.0	
4	2330393	2018-01-29 12:29:31	NaN	NaN	NaN	99.0	123.0	
5	2330393	2018-01-29	NaN	NaN	NaN	98.0	126.0	

3

```
Out[5]: MDC_PRESS_CUFF_SYS            100.000000
        MDC_PRESS_CUFF_MEAN           100.000000
        MDC_PRESS_CUFF_DIA            100.000000
        MDC_PULS_OXIM_SAT_O2           99.861496
        MDC_PULS_OXIM_PULS_RATE        99.861496
        MDC_ECG_CARD_BEAT_RATE         99.168975
        MDC_TTHOR_RESP_RATE            97.922438
        MDC_PRESS_BLD_ART_SYS          25.900277
        MDC_PRESS_BLD_ART_DIA          25.900277
        MDC_PRESS_BLD_ART_MEAN         25.900277
        MDC_ECG_V_P_C_RATE             25.484765
        MDC_TEMP                       18.005540
        MDC_ECG_AMPL_ST_II             16.897507
        MDC_PRESS_BLD_VEN_CENT_MEAN     6.371191
        MDC_ECG_AMPL_ST_I               5.401662
        MDC_ECG_AMPL_ST_III             4.432133
        MDC_PRESS_INTRA_CRAN_MEAN       2.908587
        MDC_PRESS_BLD_MEAN              2.770083
        MDC_PRESS_BLD_SYS               2.770083
```

2

Data:

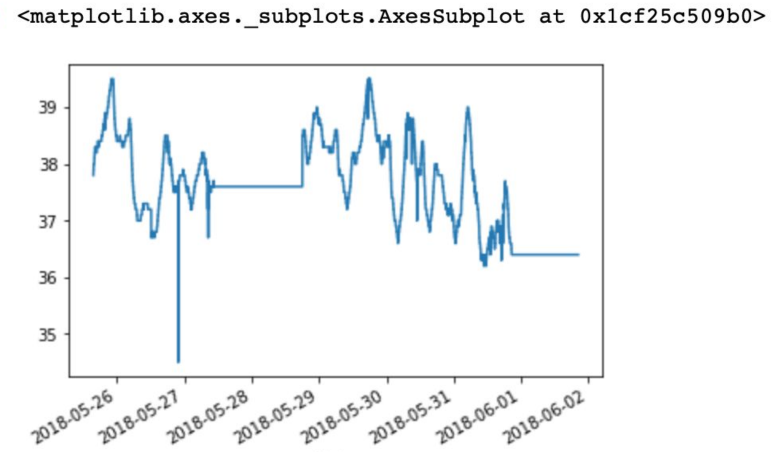
The data (fig. 3) is a high frequency physiologic ICU sensor data with 722 patients and more than 5 million observations. The total number of physiologic variables were 35 (fig. 6) out of which only 7 (fig. 7) were selected after downsampling.

35 Variables

'MDC_CO2_RESP_RATE', 'MDC_CONC_AWAY_CO2_ET',
'MDC_CONC_AWAY_CO2_INSP', 'MDC_ECG_AMPL_ST_AVF',
'MDC_ECG_AMPL_ST_AVL',
'MDC_ECG_AMPL_ST_AVR', 'MDC_ECG_AMPL_ST_I', 'MDC_ECG_AMPL_ST_II',
'MDC_ECG_AMPL_ST_III', 'MDC_ECG_AMPL_ST_V',
'MDC_ECG_CARD_BEAT_RATE',
'MDC_ECG_PACED_BEAT_RATE', 'MDC_ECG_V_P_C_RATE',
'MDC_PRESS_BLD_ART_DIA', 'MDC_PRESS_BLD_ART_MEAN',
'MDC_PRESS_BLD_ART_PULM_DIA', 'MDC_PRESS_BLD_ART_PULM_MEAN',
'MDC_PRESS_BLD_ART_PULM_SYS', 'MDC_PRESS_BLD_ART_SYS',
'MDC_PRESS_BLD_ATR_LEFT_MEAN', 'MDC_PRESS_BLD_DIA',
'MDC_PRESS_BLD_MEAN', 'MDC_PRESS_BLD_SYS',
'MDC_PRESS_BLD_VEN_CENT_MEAN', 'MDC_PRESS_CEREB_PERF',
'MDC_PRESS_CUFF_DIA', 'MDC_PRESS_CUFF_MEAN', 'MDC_PRESS_CUFF_SYS',
'MDC_PRESS_INTRA_CRAN_MEAN', 'MDC_PULS_OXIM_PULS_RATE',
'MDC_PULS_OXIM_SAT_O2', 'MDC_TEMP', 'MDC_TTHOR_RESP_RATE'

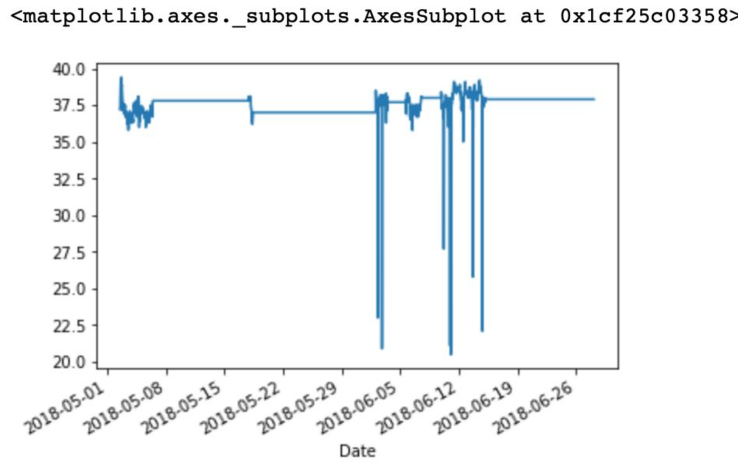
6

```
Out[241]: <matplotlib.axes._subplots.AxesSubplot at 0x1cf25c509b0>
```



5

```
Out[240]: <matplotlib.axes._subplots.AxesSubplot at 0x1cf25c03358>
```



4

7 Variables

MDC_PRESS_CUFF_SYS
MDC_PRESS_CUFF_MEAN
MDC_PRESS_CUFF_DIA
MDC_PULS_OXIM_SAT_O2
MDC_PULS_OXIM_PULS_RATE
MDC_ECG_CARD_BEAT_RATE
MDC_TTHOR_RESP_RATE

7

Methods:

The initial step is to select relevant data (>= 90% non-NaN values). The data is then filtered out in such a way that only data before the events of fever is taken.The data had two instances (cases: 16,093 and controls: 112). For cases, the data is extracted up to 6 hours back from fever event, and delta time (feverTime - recordedDateTime) is calculated respectively. Different signal processing features are extracted from 6 hour window time and the calculated delta time is used. If patients have multiple fever episodes during their stay, each episode is treated independently– provided there was 24 hour gap between them.
For controls, patients who never have temperature (fig. 4) over 38C are identified. Likewise, a random fever time is generated to calculate the respective delta time. A 6 hour window time is selected to extract significant features within that window time.

Results:

After signal processing (fig.1) using tsfresh, a total of 6,352 features (fig. 8 & 9) are selected. The next step would be to fit different machine learning models to predict the onset of fever. The accuracy of the model could be found by calculating various scores such as F1, specivity, sensitivity, area under the curve etc.

```
In [31]: extracted_features = extract_features(dfsub9_t4, column_id = "sample_id", column_sort = "delta_Reduced")
         extracted_features.to_csv("Fever_extracted_features_t4.csv")

Feature Extraction: 100%|██████████████████████████████| 90/90 [04:38<00:00,  3.10s/i
t]

In [32]: extracted_features = extract_features(dfsub9_t5, column_id = "sample_id", column_sort = "delta_Reduced")
         extracted_features.to_csv("Fever_extracted_features_t5.csv")

Feature Extraction: 100%|██████████████████████████████| 84/84 [04:26<00:00,  3.17s/i
t]
```

8

```
In [90]: extracted_features
```

variable	MDC_ECG_CARD_BEAT_RATE__abs_energy	MDC_ECG_CARD_BEAT_RATE__absolute_sum_of_changes	MDC_ECG_CARD_BEAT_RATE__agg_autocorrelation__f_
id			
2398315_1517361194.0	7916613.0	11382.0	
2398315_1517361254.0	8649.0	0.0	
2398315_1517361314.0	9216.0	0.0	
2398315_1517361374.0	18050.0	0.0	
2398315_1517361434.0	8836.0	0.0	
2398315_1517361494.0	8836.0	0.0	
2398315_1517361554.0	8649.0	0.0	
2398315_1517361614.0	8649.0	0.0	
2398315_1517361674.0	8649.0	0.0	

9

Conclusion:

The bulk of the project is data pre processing along with feature engineering of more than 5 millions data points. Different machine learning and regression models are to be applied to the extracted signal-processed features.

Acknowledgement:
Dr. Rishi Kamaleswaran, Dr. Akram Mohammed, The University of Tennessee Health Science Center
Prof. Kathryn Montovan, Andrew McIntyre, Bennington College

References:
https://towardsdatascience.com/dengue-fever-and-how-to-predict-it-a32eab1db