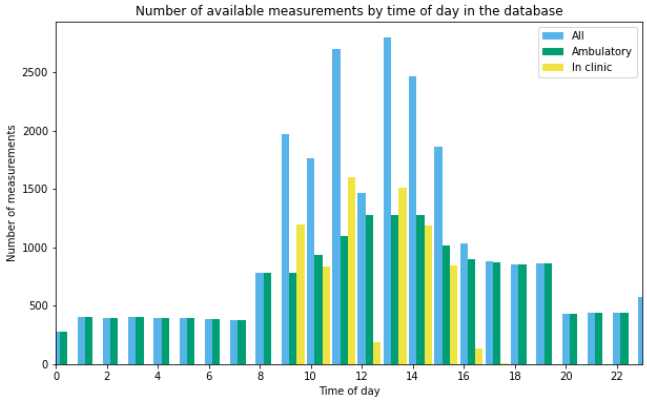
Supplementary Material:

Robust Feature Selection for Continuous BP Estimation in Multiple Populations: Towards Cuffless Ambulatory BP Monitoring

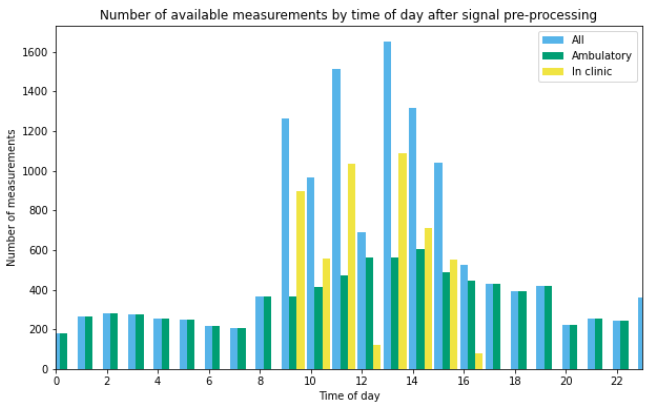
Ana Cisnal, *Student Member, IEEE*, Yanke Li, Bertram Fuchs, Mehdi Ejtehadi, Robert Riener, *Senior Member, IEEE*, and Diego Paez-Granados[†], *Member, IEEE*

Available data after and before signal pre-processing

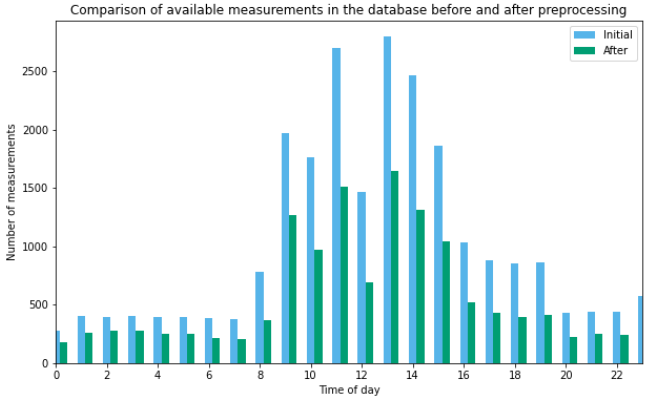
The AuroraBP database exclusively contains PPG signal segments with a reasonably acceptable quality index. Consequently, signals with significant noise were excluded and not included in the public database. Moreover, it is important to note that during the day, the ambulatory blood pressure (BP) device triggers automatically every 30 minutes, whereas during the night, it triggers every 60 minutes. Additionally, during the day, clinical tests are conducted, comprising 14 additional tests, including various arm positions (sitting arm down, sitting arm lap, sitting arm up), supine, standing, walking, running, and cool down exercises. Therefore, despite the PPG signals being significantly noisier during the day, there is a greater availability of data during daytime hours. In the following figure, the number of available measurements of the database are shown:



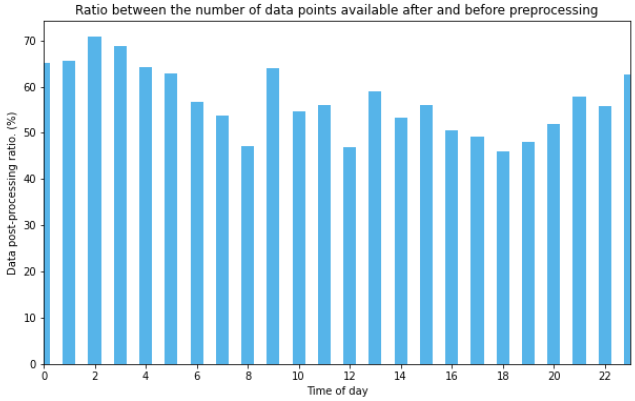
After the signal pre-processing, some signal segments are entirely discarded because they do not meet the quality standards of the proposed method. The number of available measurements by time of day after signal-preprocessing is shown in the following figure:



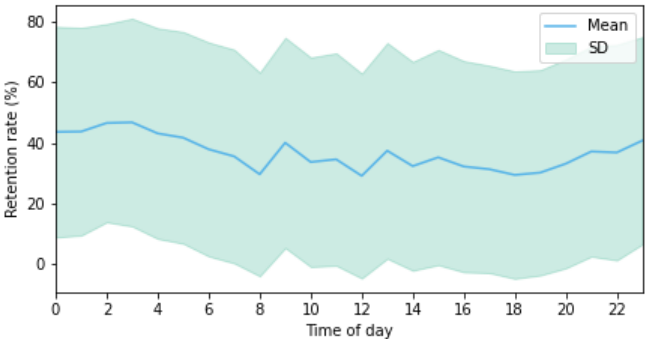
We can see the comparison of available measurements before and after the signal pre-processing:



The data post-processing ratio, which is the between the number of data points available after preprocessing and before preprocessing, expressed as a percentage, is 56.97 (7.22) %. The evolution during time of day of the data post-processing ratio is shown in the following figure:



While the data post-processing ratio is slightly lower during the day, there is no significant difference between daytime and nighttime hours. It can be observed that there are more signal segments during the day than during the night in the original data, and this higher quantity of data during the day persists after signal preprocessing. Additionally, the retention rate of the available measurements, which is the ratio between the final time after pre-processing and the total time of the signal segment expressed in percentage, is calculated. The signal retention rate after pre-processing was 35.22 (35.00) %, with distinct rates of 32.54 (34.63) % for in-clinic measurements and 41.25 (35.06) % for ambulatory measurements.



The significance of this data post-processing ratio lies in its crucial role within the proposed method, where each feature's value within a signal segment is averaged to serve as input for the machine learning models. While the retention rate, indicating the length available from a signal segment, provides information about the signal quality, the primary focus is on the data post-processing ratio. This ratio serves as a key indicator of the quantity of feature vectors that can be effectively inputted into the model.