Supplementary Material 1:

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*

Feature Space from PPG signals

The generated features can be divided into four main groups: time domain (TF), frequency-based (FF), statistical (SF), and demographic features (DF)

1) Time-domain Features (TF)

The proposed time-domain features can be divided into six subcategories regarding intensity, time, area, slope, branch width, and others. Several of these features are computed using one or two fiducial points. The fiducial points include the onset (O), valley (V), systolic peak (S), maximum derivative point (MD), diastolic peak (D), inflection point (IP), dicrotic notch (DN), point a, point b, and intersection point (IT).

Abbreviation	Name	Definition
I - Intensity-based features		
i_{x1}	Intensity	Absolute Intensity or amplitude of point x1
i_{x1-x2}	Relative intensity	Relative intensity or amplitude between point x1 and x2 $I_{X1-X2}=I_{x1}- I_{x2}$
ir_{X1-X2}	Intensity ratio	Intensity ratio of points x1 and x2 $RI_{X1-X2}=I_{x1}/ I_{x2}$
ni_{X}	Normalized intensity	Normalized intensity within the range of 0 to 1
ni_{X1-X2}	Relative normalized intensity	Relative normalized intensity of amplitude between point x1 and x2 $ni_{X1-X2}= ni_{X1-} ni_{X2}$
id_{x1}	Intensity of the first derivative	Absolute intensity or amplitude of the first derivative at point x1.
id_{x1-x2}	Relative intensity of the first derivative	Relative intensity or amplitude between point x1 and x2 in the first derivative
$id2_{x1}$	Intensity of the second derivative	Absolute intensity or amplitude of the second derivative at point x1.
ird2x1-x2	Relative intensity of the second derivative	Relative intensity or amplitude between point x1 and x2 in the first derivative
Im	Mean intensity	Mean intensity of the waveform
mean_irx1	Mean intensity ratio	Mean intensity ratio when dividing the waveform into two segments at point x1.
T - Time-based features		
t_{x1}	Relative time	Time between key-point x1 and v Note that t_{v1} is the total duration of the pulse waveform
t_{x1-x2}	Time difference	Time difference between points x1 and x2 $t_{x1-x2} = t_{x1} - t_{x2}$

tn_{x1}	Normalized time	Normalized time from the onset to the x1 point with respect to the total duration of the pulse waveform. $nt_{x1} = t_{x1}/t_{v1}$
tr_{x1-x2}	Time ratio	Ratio between the time from the onset to x1 and the time from the onset to x2. $tr_{x1-x2} = t_{x1} / t_{x2}$
trn_{x1-x2}	Normalized time ratio	Normalized time ratio between points $trn_{x1-x2} = t_{x1-x2} / t_{v1}$
A - Area-Related features		
A_{x1-x2}	Area	Area under the curve defined by points x_1 and x_2
ART_{x1-x2}	Total area ratio	Total area ratio between S and D. Ratio of the area under the curve between x_1 and x_2 to the total area under the pulse waveform. $ART_{x1-x2} = A_{x1-x2}/_{Av-v1}$
ARs_x	Ratio of areas	Break the area under PPG into two parts based on the key point x, where A_1 and A_2 are the areas of the first and second parts. $RA= A_{V-x} /A_{x-v1}$
SLP – Slope		
SLP_{X1}	Slope	Slope from v to key-point x1 $SLP_{X1} = (I_{x1}-I_V) / t_{x1}$
SLP_{x1-n}	Normalized slope	Slope from v to key-point x1 $nSLP_{X1} = (I_{x1}-I_V) / tn_{x1}$
BW - Branch-width related features x = 10%, 25%, 33%, 50%, 66%, 75% and 90%		
SBW_x	Systolic Branch Width	Systolic branch width at x% of pulse height of PPG
DBW_X	Diastolic Branch Width	Diastolic branch width at x% of pulse height of PPG
BW_x	Branch Width	Branch width at x% of the pulse height of PPG $BW_x = SBW_X + DBW_X$
BWR_x	Branch Width Ratio	Branch width ratio at x% of the pulse height of PPG $BWR_x = DBW_X/SBW_X$
O - Others		
corr	Autocorrelation	Autocorrelation
centr	Centroid	Centroid
min_p	Minimum peaks	Number of local minima of the signal
max_p	Maximum peaks	Number of local maxima of the signal
mean_abs_diff	Mean absolute differences	Mean absolute differences of the signal
mean_diff	Mean of differences	Mean of differences of the signal
med_abs_diff	Median absolute differences	Median absolute differences of the signal
med_diff	Median of differences	Median of differences of the signal
dist	Distance	Signal traveled distance using the hypotenuse between 2 points
sadif	Sum of absolute differences	Sum of absolute differences of the signal
zc1d	Zero cross First Derivative	Number of times that the first derivative cross zero
zc2d	Zero cross Second Derivative	Number of times that the second derivative cross zero
zc3d	Zero cross Third Derivative	Number of times that the third derivative cross zero
total_ene	Total energy	Total energy
slope	Slope	Slope of the signal
abs_ene	Absolute energy	Absolute energy of the signal
ent_kde	Entropy	Entropy calculated by KDE (Kernel Density Estimate))
ent_gauss	Entropy	Entropy calculated by Gauus

$coef_x$	Polynomial coefficient x	The coefficients of a 15th-order polynomial are employed to accurately fit the PPG pulse waveform.
RI or AI	Reflexion index	$RI = I_{IP}/I_S$
LASI	Large Artery Stiffness Index	$LASI = h/\Delta T_{S-I}$, where H is the height of the subject ($H \approx 1$).
mNPV	Normalized Pulse Volume	$mNPV = I_{ac} / (I_{ac} + I_{dc})$, where I_{ac} is the peak-to-peak amplitude and I_{dc} is the average of the pulse
PPGK	characteristic value or K value	$PPGK = (I_m - I_v)/(I_s - I_v)$ respect to the baseline

2) Frequency-based features (FT)

These features were derived from the beat-to-beat PPG pulse waveforms and PPG segments, which were constructed exclusively using validated PPG pulses (i.e., removing invalid pulses from the original 30-second segment during the signal processing). Features labeled with an asterisk (*) denote extraction from PPG segments, whereas those without an asterisk were derived from pulse waveforms.

Ab	Code Name	Name	Definition
fsqi	fsqi	Relative power	Ratio between two frequency bands. The frequency bandwidth for the ratios' numerator is from 1 to 2.25 Hz and ratios' denominator is from 0 to 8 Hz
fsqi1	fsqi1	Relative power at the first harmonic	Ratio between two frequency bands. The frequency bandwidth for the ratios' numerator is from 0 to frequency of first harmonic and ratios' denominator is the whole spectrum
fsqi2	fsqi2	Relative power at the second harmonic	Ratio between two frequency bands. The frequency bandwidth for the ratios' numerator is from 0 to frequency of second harmonic and ratios' denominator is the whole spectrum
fsqi3	fsqi3	Relative power at the third harmonic	Ratio between two frequency bands. The frequency bandwidth for the ratios' numerator is from 0 to frequency of third harmonic and ratios' denominator is the whole spectrum
f1	f1	Frequency of the first harmonic	Predominant frequency of the signal
mag_{f1}	mag_f1	Magnitude of the first harmonic	Magnitude of the predominant frequency of the signal
f2	f2	Frequency of the second harmonic	Frequency of the second harmonic of the signal
mag_{f2}	mag_f2	Magnitude of the second harmonic	Magnitude of the second harmonic of the signal
f3	f3	Frequency of the third harmonic	Frequency of the third harmonic of the signal
mag_{f3}	mag_f3	Magnitude of the third harmonic	Magnitude of the third harmonic of the signal
Sp. distance	spectral_distance	Spectral distance	Distance of the signal's cumulative sum of the FFT elements to the respective linear regression.
Fundamental f	fundamental_frequency	Fundamental frequency	Predominant frequency of the signal
Max PS	max_power_spectrum	Maximum power spectrum	Maximum value of the power spectrum density

Max f	max_frequency	Maximum frequency	Maximum frequency of the signal.
Med f	median_frequency	Median frequency	Median frequency of the signal.
Sp. centroid	spectral_centroid	Spectral centroid	Barycenter of the spectrum
Sp. decrease	spectral_decrease	Spectral decrease	Represents the amount of decreasing of the spectra amplitude.
Sp. K	spectral_kurtosis	Spectral kurtosis	Measures the flatness of a distribution around its mean value.
Sp. S	spectral_skewness	Spectral skewness	Measures the asymmetry of a distribution around its mean value.
Sp. spread	spectral_spread	Spectral spread	Measures the spread of the spectrum around its mean value.
Sp. SLP	spectral_slope	Spectral slope	Computes the spectral slope.
Sp. variation	spectral_variation	Spectral variation	Computes the amount of variation of the spectrum along time. Spectral variation is computed from the normalized cross-correlation between two consecutive amplitude spectra.
Sp. Max Peaks	spectral_maxpeaks	Spectral maxpeaks	Number of maximum spectral peaks of the signal.
Sp. Roll-off	spectral_roll_off	Spectral Roll-off	The spectral roll-off corresponds to the frequency where 95% of the signal magnitude is contained below of this value.
Sp. Roll-on	spectral_roll_on	Spectral Roll-on	The spectral roll-on corresponds to the frequency where 5% of the signal magnitude is contained below of this value.
HRER	human_range_energy	Human range energy ratio	The human range energy ratio is given by the ratio between the energy in frequency 0.6-2.5Hz and the whole energy band.
PW	power_bandwidth	Power spectrum density bandwidth	It corresponds to the width of the frequency band in which 95% of its power is located.
SE	spectral_entropy	Spectral entropy	Computes the spectral entropy of the signal based on Fourier transform.
WE	wavelet_entropy	Wavelet entropy	Computes CWT Shannon entropy of the signal.

3) Statistical features (SF)

Ab.	**Name/Definition**
s	Skewness
k	Kurtosis
mav	Mean Absolute Value
median	Median
mad	Mean Absolute Deviation
med_ad	Median Absolute Deviation
rms	Root-Mean-Square
sd	Standard Deviation
sf	Shape Factor
if	Impulse Factor
cf	Crest Factor

v	Variance
irq	Interquartile range
p	Perfusion

4) Demographic features (DF)

Abbreviation	**Name/Definition**
age	Age
weight	Weight
bmi	Body Mass Index
baseline_sbp	Systolic Blood Pressure at rest
baseline_dbp	Diastolic Blood pressure at rest