Third Stakeholder Workshop

Classification of Skin Tone Using PPG Signals

Urs Hackstein (THM)
Joint work with Philip Aston (NPL), Alen Bosnjakovic (IMBiH), Padmini Krishnadas (NPL)

14 April 2026

PPG signals and Skin Tone

More details:

Krishnadas, P.; Hackstein, U.; Bosnjakovic, A.; Aston, P.J. Fuzzy accuracy compensates for label subjectivity in classification of skin tone using wearable photoplethysmography signals. IEEE International Conference on Fuzzy Systems, 2026.

PPG signals and Skin Tone

- QUMPHY investigates various biases with influence on the evaluation of ML methods on PPG signals
- Example: Skin Tone
- Cf. Independent Review in the UK: *Equity in Medical Devices: Independent Review* (https://www.gov.uk/government/publications/equity-in-medical-devices-independent-review-final-report, 2024)

➔ "extensive evidence of poorer performance of pulse oximeters for patients with darker skin tone"

Problem description

Question:

Is it possible to determine the skin tone of a subject from the PPG signal?

-> If skin tone had no influence on PPG signals, the labels would be allocated randomly, i.e. approx. 1 in N correct for Multiclass classification with N classes

Fitzpatrick scale

- Thomas Fitzpatrick (US-dermatologist, 1975):
  Six classes:
  1: light ….. 6: dark
- Original Goal: Determine the correct UV dose for treating skin disease

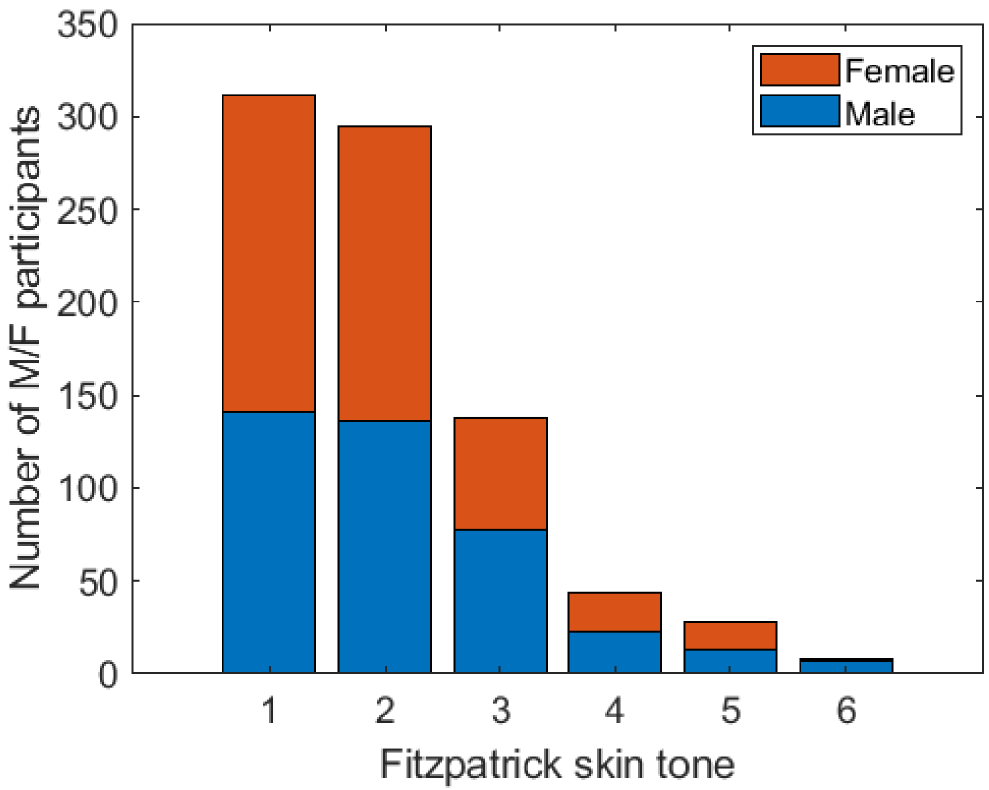
- SEVERE LIMITATION: Fitzpatrick scale is purely subjective, not based on any measurement
- Labels may differ by one class from the “correct“ class

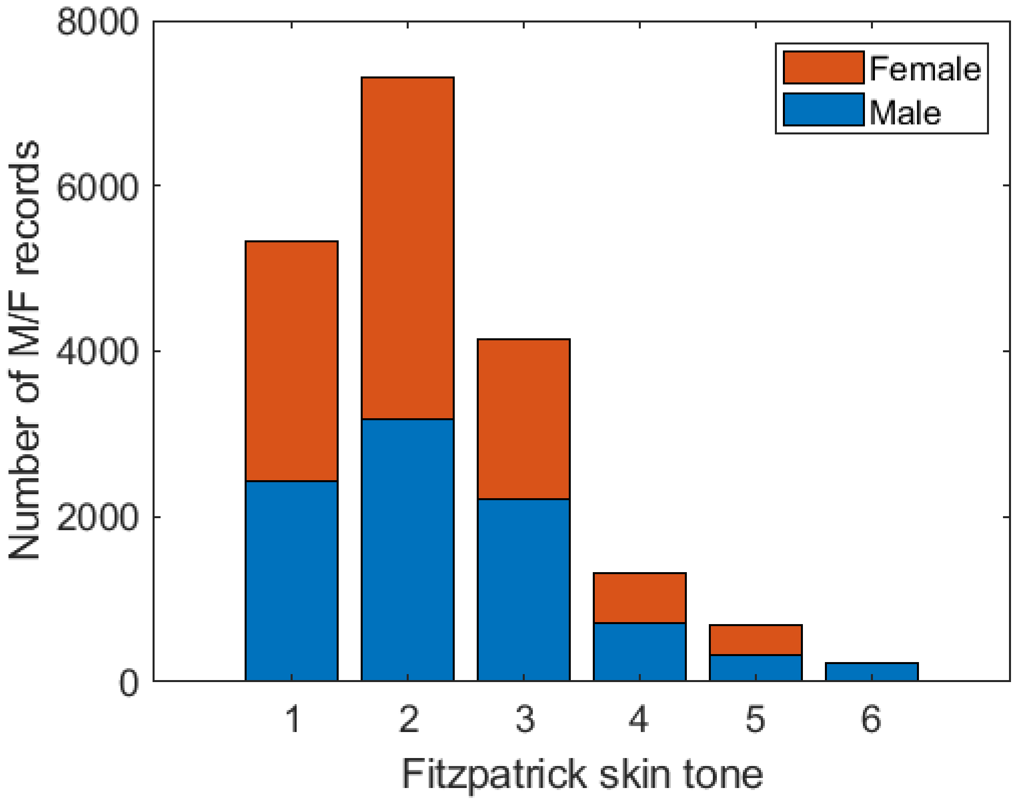
Data

- AuroraBP:
  - available on request from Microsoft
  - PPG waveforms and Fitzpatrick scale labels for 823 subjects (51.6% female)
  - collected at 250Hz, resampled to 500Hz
  - auscultatory and oscillometric measurement protocols
  - combined to 19,017 records

- Work with given quality measures:
  Optical quality ≥ 0.65

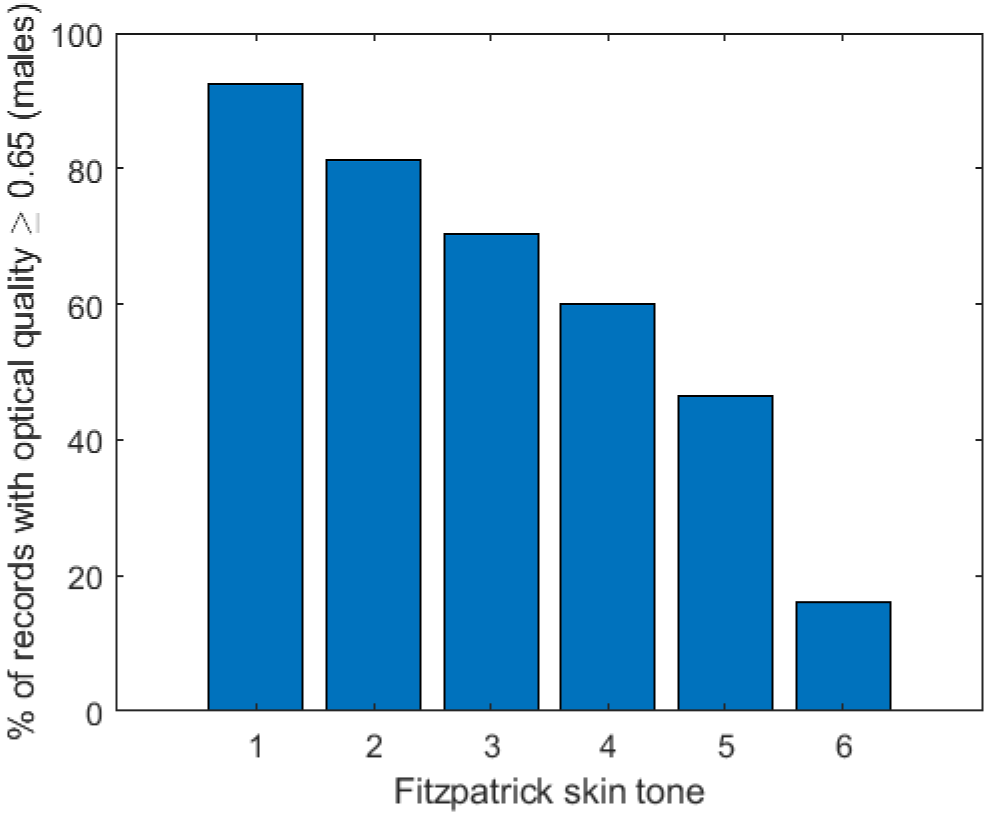
Data

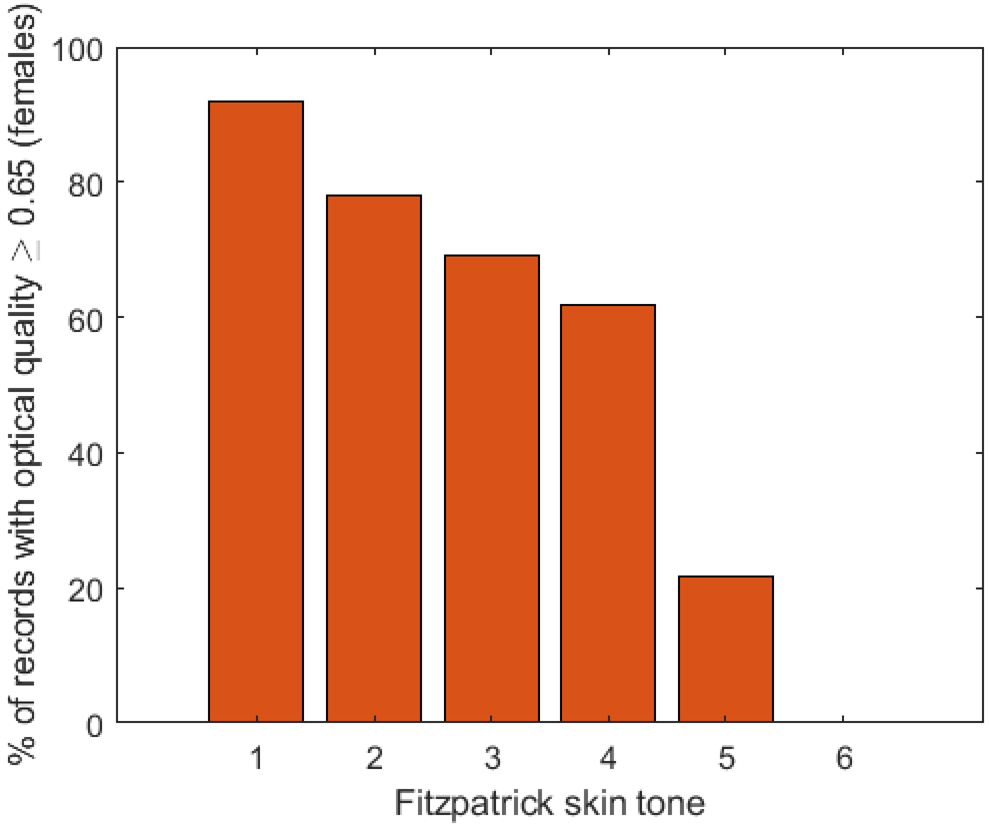




Imbalanced with respect to number of participants and also records

Data





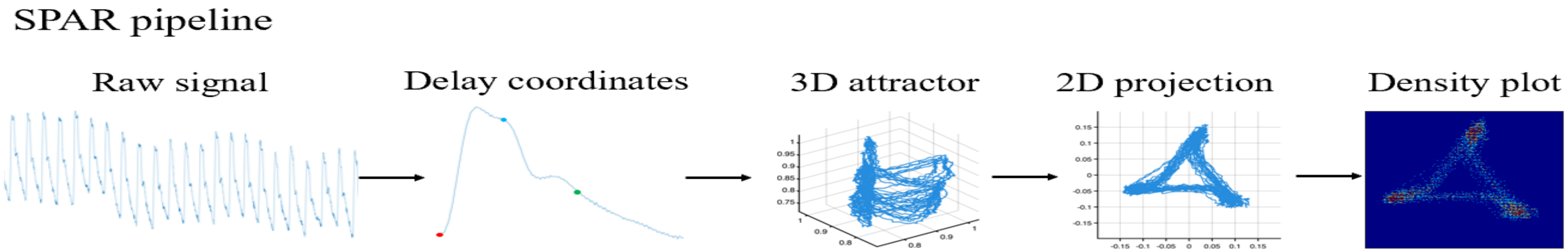
Waveforms have poorer quality for darker skin tone

Methods

Three different classification approaches:

1) Machine and Deep Learning on raw or just filtered signals
2) ML and DL with Interpretable PPG features using PulseAnalyse code (cf. P.H. Charlton)
3) Deep Learning with SPAR attractor images

SPAR =Symmetric Projection Attractor Reconstruction (due to Philip Aston and Manasi Nandi)
-> transforms signals into images
-> method that can extract additional information pertaining to the morphology and variability of physiological waveforms



Results

- Stratified 5-fold cross validation per subject, all records with optical quality $\geq\ 0.65$

- Best results for our three approaches:

Method	Classifier	Loss function	accuracy
Raw signals	Gradient Boosted Decision Tree (ydf package)	Multinomial log likelihood	55.03%
Interpretable features	Neural Network (Matlab)	Cross entropy	42.19%
SPAR images	ResNet 152	Cross entropy	44.39%

Results: Fuzzy accuracy

- **Observation**: Fitzpatrick scale is subjective, labels may differ from the „correct“ label by 1 class

- **Solution**: “Fuzzy accuracy“

- Predicted class is correct iff $|\text{pred. class} - \text{labelled class}| \leq 1$
- For example: labelled class 2 -> predicted classes 1,2,3 are counted as correct

Results: Fuzzy accuracy

Best results for the three different methods

Method	Classifier	Loss function	Accuracy	Fuzzy accuracy
Raw signals	Gradient boosted decision tree	Fuzzy cross entropy	53.36%	95.94%
Interpretable PPG feaures	Neural Network (Matlab)	Cross entropy	42.19%	87.05%
SPAR images	VGG16	Earth Mover Distance	40.47%	86.00%

Conclusion

- Our results prove the impact of skin tone on PPG signals
- Skin tone should be taken in account within the analysis of PPG signals

www.qumphy.ptb.de

gitlab.com/qumphy

qumphy-software.rtfd.io