3rd Stakeholder Workshop

22HLT01 - QUMPHY

Quantifying uncertainties for machine learning with medical applications

Nando Hegemann

Project Overview

Title

Uncertainty quantification for machine learning models applied to photoplethysmography signals

Infos

Budget: 2.1 M€ **Funding Period:** 07/2023 - 06/2026 **Coordination:** PTB

KERs

- Benchmark problems & datasets
- Good practice guide
- Software repository & user interface

| Define good practice for UQ of ML in medical applications
| Focus on end-user applicability

Team

- 16 consortium members
- 9 countries
- 30+ people
- 1 external ethics advisor

Focus, Impact & Outreach


Noninvasive wearable devices for PPG signal collection

PPG signals

FAIRness

- Investigate influence of **sex and skin tone** on predictions and uncertainties of ML models when applied to PPG signals
- Ensure **transparency and public accessibility** of benchmarks, good practice guide and software repository

Excellence

- **Expert consortium** to design, develop and disseminate solutions to increase credibility of ML in medical applications
- Constant **feedback loop** with target communities to ensure relevance of project output Impact

Impact

- Establish guidelines to quantify uncertainties of ML in medical diagnostics **affecting billions of people**
- **Exploitation** of results by medical device, digital, healthcare, scientific and metrological communities

Implementation

- Modularity of WPs allows **efficient and adaptable realization** of the project to minimize overall risk
- **Combination of multiple perspectives** (metrological, scientific, economic and clinical) through a broad consortium

Project Need and Aim

- Investigate performance of machine learning (ML) for photoplethysmography (PPG) signals on **diagnosis of diseases** (elevated blood pressure, diabetes, atrial fibrillation)
- Define standard **benchmarks** to compare ML applications
- Increase trust in diagnostic predictions by defining **uncertainty quantification (UQ) measures** to assess ML algorithms
- Improve reproducibility and reliability of ML with open **software repository and guide** on uncertainty quantification
- Encourage **feedback and uptake** of benchmarks and testing framework by medical device, digital and healthcare communities

Project Structure

WP1 – Uncertainty Quantification

Focus

1. Develop UQ methods and ML algorithms and apply them to 3 classification/regression tasks involving PPG signals
2. Validate uncertainties of ML models on benchmark problem

Progress beyond the State-of-the-Art

1. Develop an open-source software repository ready to use by digital healthcare companies
2. Document ML models and UQ methods to contribute to a good practice guide on UQ for ML algorithms applied to PPG signals

WP2 – Data and Promotion

Focus

1. Generate at least 5 benchmark problems and make them publicly available
2. Engage with medical device, digital and healthcare communities
3. Develop tools for target end-users to employ UQ for ML applied to tasks involving PPG data

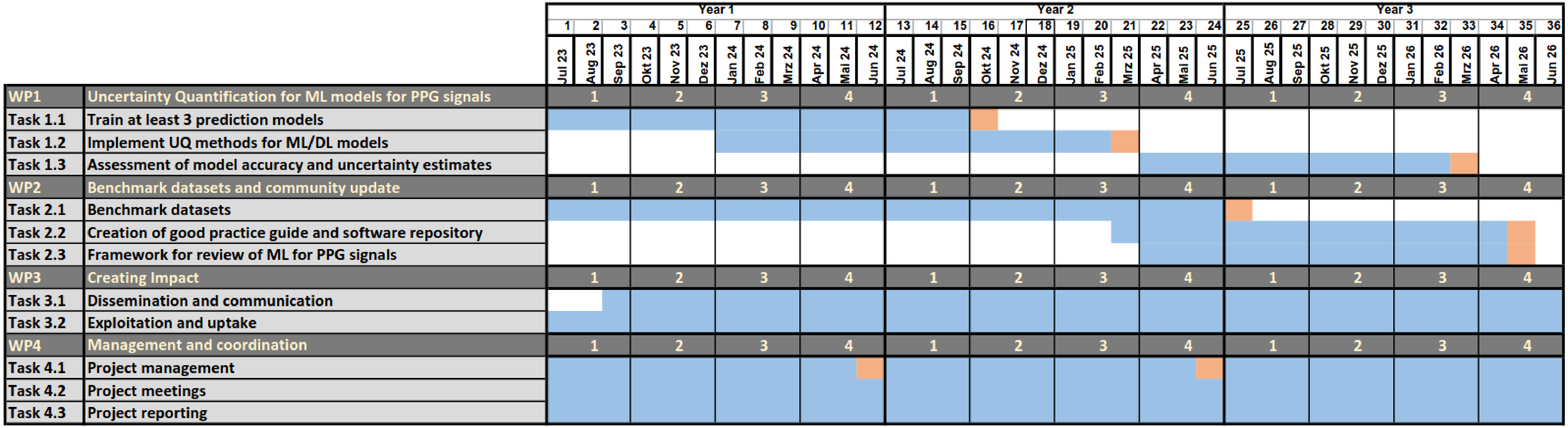
Progress beyond the State-of-the-Art

1. Design industry relevant benchmarks and make them publicly available
2. publish a good practice guide describing common ML algorithms, UQ methods, benchmarks and examples
3. Engage with medical device, digital and healthcare communities for constant feedback on end-user relevance

WP3 – Impact and Outreach

Disseminate results to target communities to increase uptake of the **benchmark datasets**, **good practice guide** and **software repository** as a basis of UQ for ML applied to PPG signals

Project Progress



External Ethics Review: Dr. Jenny Venton, The Royal Surrey NHS Foundation Trust

- Ethics self-assessment
- Feedback and recommendations
- Discussion of potential misuse of results
- Active involvement in project pathway
- Continuous implementation oversight

D7 First report from the ethics advisor to 22HLT01

D8 Second report from the ethics advisor to 22HLT01

Deliverables

D1: Report on **developing** methods based on existing supervised **machine learning and/or deep learning models** using PPG signals representing at least **3 classification** and **3 regression** uncertainty determinations, including a **comparison of model performance** for one prototypical regression and one prototypical classification problem (**Publication:** https://doi.org/10.1016/j.bspc.2026.109831)

D2: Report on developing methods on **uncertainty quantification for** one prototypical **machine learning and/or deep learning** classification and one prototypical regression problem using PPG signals including the effects of both aleatoric (data) uncertainty and epistemic (model) uncertainty on model predictions (**Publication:** https://doi.org/10.1088/3049-477X/ae4c8e)

D3: Report on **validating the uncertainties** obtained for existing machine learning and deep learning models from D1 and D2. This includes the comparison of accuracy and uncertainty of at least 3 classification and 3 regression machine learning and/or deep learning models in order to identify models and methods which have high accuracy and low uncertainty for a wide range of tasks (**under preparation**)

D4: Report that at least 5 **datasets** using real and/or synthetic PPG data have been generated, that can be used **to benchmark accuracy and uncertainty** of supervised machine learning and deep learning models. This includes making the 5 reference problems and their respective 5 datasets available to the medical device and digital health communities (**Report:** https://arxiv.org/abs/2604.01398)

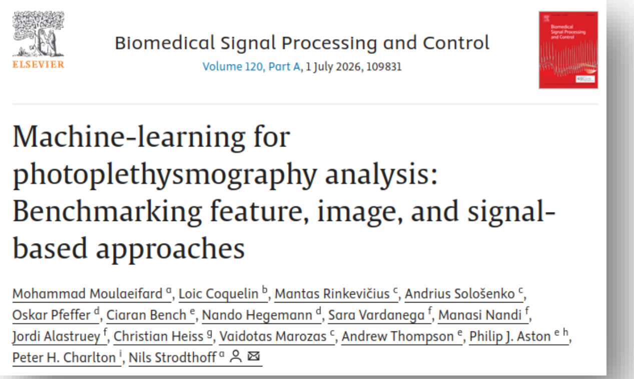
D5: Good practice guide for quantifying uncertainties for machine learning models applied to PPG signals. This includes an email confirmation to EURAMET that a repository containing the developed software has been made publicly available (**under preparation**)

D6: Report containing guidelines to support the adoption of the benchmarking problems and the development of a **framework for independent review** of the proposed machine learning models by e.g., medical device, digital and health communities to assist them in getting regulatory approval (**Software:** gitlab.com/qumphy/qumphy-software, **Documentation:** qumphy-software.rtfd.io)

Articles & Good Practice

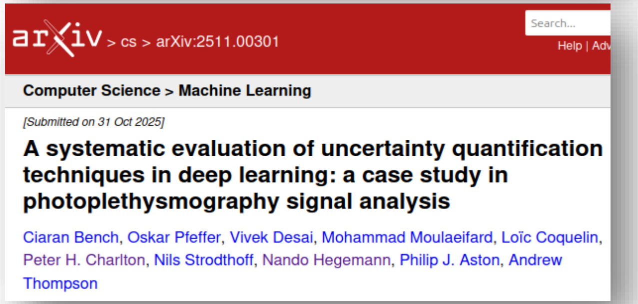
Articles (Peer Reviewed)

1. AD Rodway, L Hanna, J Harris, R Jarrett, C Allan, F Pazos Casal, BCT Field, MB Whyte, N Ntagiantas, I Walton, A Pankhania, SS Skene, GD Maytham, C Heiss, "**Prognostic and predictive value of ultrasound-based estimated ankle brachial pressure index at early follow-up after endovascular revascularization of chronic limb-threatening ischaemia: a prospective, single-centre, service evaluation**", eClinicalMedicine 68(102410), 2024
   published: https://doi.org/10.1016/j.eclinm.2023.102410
2. M Rinkevičius, J Lazaro, E Gil, P Laguna, PH Charlton, R Bailon, V Marozas, "**Obstructive Sleep Apnea Characterization: A Multimodal Cross-Recurrence-Based Approach for Investigating Atrial Fibrillation**", IEEE Journal of Biomedical and Health Informatics, 2024
   published: https://doi.org/10.1109/JBHI.2024.3428845
3. L Hannab, AD Rodwaya, P Garchac, L Maynardc, J Sivayogic, O Schlagere, J Madaricf, V Boch, L Buschg, MB Whytec, SS Skenec, J Harrisi, C Heiss, "**Safety and procedural success of daycase-based endovascular procedures in lower extremity arteries of patients with peripheral artery disease: a systematic review and meta-analysis**", eClinicalMedicine 75(102788), 2024
   published: https://doi.org/10.1016/j.eclinm.2024.102788
4. P Charlton, V Marozas, E Mejía-Mejía, PA Kyriacou, J Mant, "**Determinants of photoplethysmography signal quality at the wrist**", PLOS Digital Health 6(e0000585), 2025
   published: http://dx.doi.org/10.1371/journal.pdig.0000585
5. C Teichert, U Hackstein, T Krüger, S Bernhard, "**Noninvasive detection of lower extremity artery disease using multi-site photoplethysmographic signals and machine learning**", npj Biosensing 1(2), 2025
   published: https://doi.org/10.1038/s44328-025-00044-z
6. M Moulaeifard, PH Charlton, N Strodthoff, "**Generalizable deep learning for photoplethysmography-based blood pressure estimation - A Benchmarking Study**", Machine Learning: Health, 2025
   published: http://dx.doi.org/10.1088/3049-477X/ae01a8
7. C Bench, V Desai, M Moulaeifard, N Strodthoff, P Aston, A Thompson, "**Uncertainty quantification with approximate variational learning for wearable photoplethysmography prediction tasks**", *Mach. Learn.: Health* **1** 015013, 2025
   published: https://www.doi.org/10.1088/3049-477X/ae0b74
8. M Moulaeifard, L Coquelin, M Rinkevičius, A Sološenko, O Pfeffer, C Bench, N Hegemann, S Vardanega, M Nandi, J Alastruey, C Heiss, V Marozas, A Thompson, PJ Aston, PH Charlton, N Strodthoff, "**Machine-learning for photoplethysmography analysis: Benchmarking feature, image, and signal-based approaches**", Biomedical Signal Processing and Control 120(A), 2026
   published: https://doi.org/10.1016/j.bspc.2026.109831
9. C Bench, O Pfeffer, V Desai, M Moulaeifard, L Coquelin, PH Charlton, N Strodthoff, N Hegemann, PJ Aston, A Thompson, "**A systematic evaluation of uncertainty quantification techniques in deep learning: a case study in photoplethysmography signal analysis**", preprint available
   preprint: https://arxiv.org/abs/2511.00301


Biomedical Signal Processing and Control
Volume 120, Part A, 1 July 2026, 109831

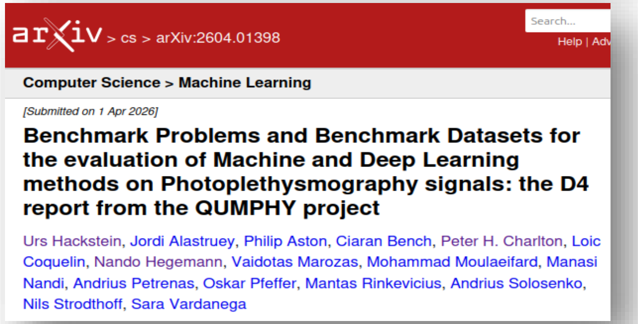
Machine-learning for photoplethysmography analysis: Benchmarking feature, image, and signal-based approaches

Mohammad Moulaeifard [a], Loic Coquelin [b], Mantas Rinkevičius [c], Andrius Sološenko [c], Oskar Pfeffer [d], Ciaran Bench [e], Nando Hegemann [d], Sara Vardanega [f], Manasi Nandi [f], Jordi Alastruey [f], Christian Heiss [g], Vaidotas Marozas [c], Andrew Thompson [e], Philip J. Aston [e h], Peter H. Charlton [i], Nils Strodthoff [a]


arXiv > cs > arXiv:2511.00301
Computer Science > Machine Learning
[Submitted on 31 Oct 2025]

A systematic evaluation of uncertainty quantification techniques in deep learning: a case study in photoplethysmography signal analysis

Ciaran Bench, Oskar Pfeffer, Vivek Desai, Mohammad Moulaeifard, Loïc Coquelin, Peter H. Charlton, Nils Strodthoff, Nando Hegemann, Philip J. Aston, Andrew Thompson


arXiv > cs > arXiv:2604.01398
Computer Science > Machine Learning
[Submitted on 1 Apr 2026]

Benchmark Problems and Benchmark Datasets for the evaluation of Machine and Deep Learning methods on Photoplethysmography signals: the D4 report from the QUMPHY project

Urs Hackstein, Jordi Alastruey, Philip Aston, Ciaran Bench, Peter H. Charlton, Loic Coquelin, Nando Hegemann, Vaidotas Marozas, Mohammad Moulaeifard, Manasi Nandi, Andrius Petrenas, Oskar Pfeffer, Mantas Rinkevicius, Andrius Solosenko, Nils Strodthoff, Sara Vardanega

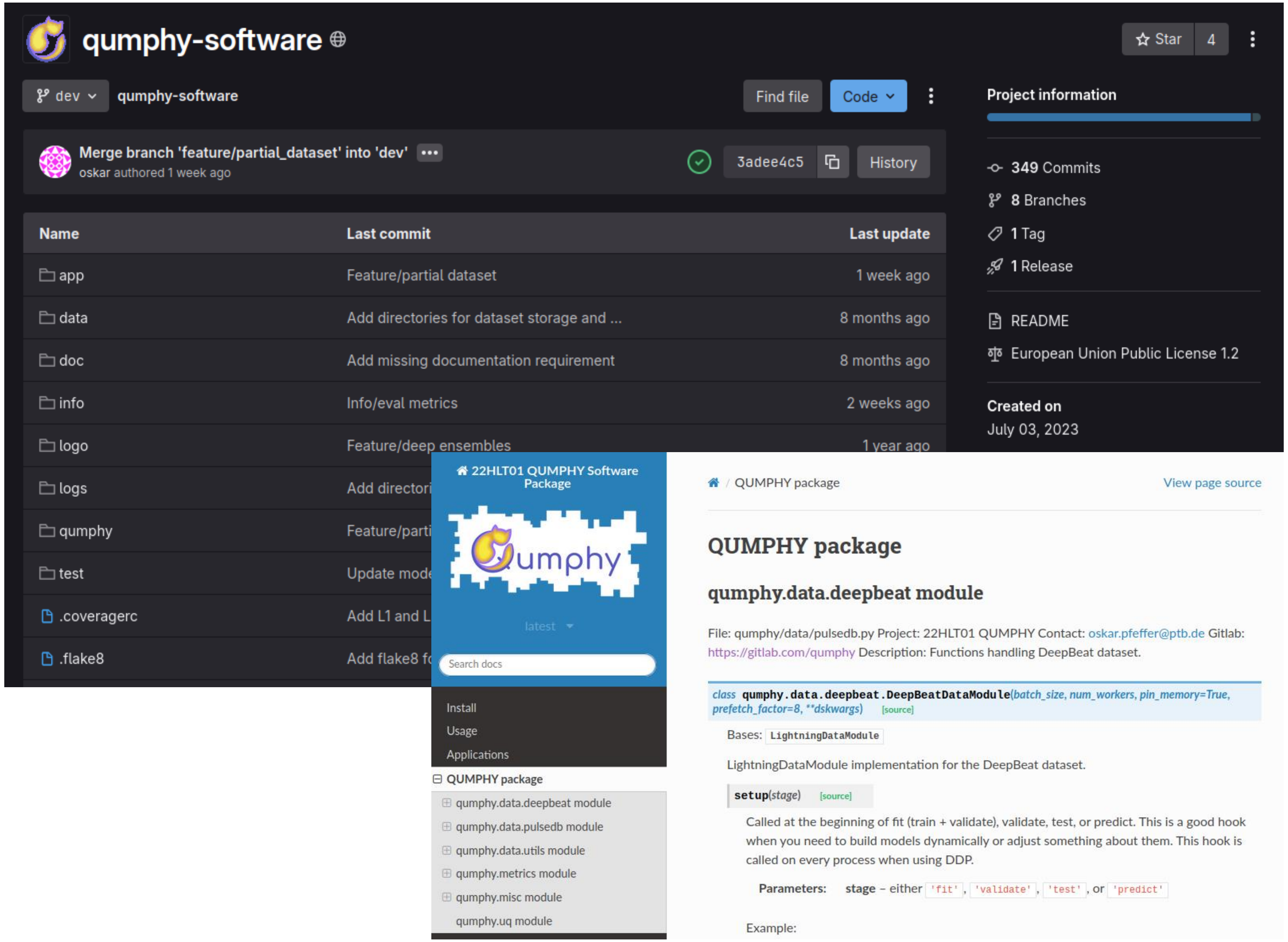
Software Repository

Content (wip)

- ML model architectures
- UQ methods
- Data download / processing
- Evaluation scripts
- Tutorials

Focus

- Modularised architecture
- Customisability through config files
- Minimal package dependencies
- Comes with setup environment
- Automatic formatting & testing
- Version control
- Automatic & public documentation
- Reproducibility of results (scripts)



Additional Software

D1 code repository: gitlab.com/qumphy/d1-code

README.md

Machine-learning for Photoplethysmography Analysis:

Benchmarking Feature, Image, and Signal-Based Approaches

This repository contains the official code for the study presented in the paper:

"Machine-learning for Photoplethysmography Analysis: Benchmarking Feature, Image, and Signal-Based Approaches"
Mohammad Moulaeifard, Loic Coquelin, Mantas Rinkevičius, Andrius Sološenko, Oskar Pfeffer, Ciaran Bench, Nando Hegemann, Sara Vardanega, Manasi Nandi, Jordi Alastruey, Christian Heiss, Vaidotas Marozas, Andrew Thompson, Philip J. Aston, Peter H. Charlton, Nils Strodthoff, arXiv:2502.19949 (2025) (https://arxiv.org/abs/2502.19949)

📌 Overview

This project investigates various machine learning strategies for analyzing photoplethysmography (PPG) signals. Two key clinical applications are explored:

- **Cuffless Blood Pressure (BP) Estimation:** Predicting systolic and diastolic blood pressure from raw PPG data.
- **Atrial Fibrillation (AF) Detection:** Classifying PPG recordings to identify the presence of atrial fibrillation.

We benchmark three types of input representations:

- **Raw Time Series:** Direct analysis of PPG signals using deep neural networks.
- **Feature-Based:** Using clinically interpretable features extracted from PPG waveforms.
- **Image-Based:** Converting PPG signals into time-frequency representations (e.g., scalograms) and analyzing them with CNNs.

🛠️ Repository Structure

1. Data Preprocessing

The `Preprocessing/` folder contains scripts for preprocessing precedure. This project utilizes the VitalDB dataset which is part of **PulseDB dataset** (Wang et al., 2023) for **BP estimation** and **Deepbeat dataset** (Torres-Soto J & Ashley EA, 2020) for **AF detection** tasks.

📌 **To prepare the dataset**, follow the instructions in the **PulseDB_Deepbeat_preprocessing** section of this repository:
🔗 Preprocessing

2. Model Training

The `Processing/` folder contains scripts to train a variety of models across the three input modalities.

Training Pipelines

Raw Time Series Models (T)

- **BAsedLine, XResNet1d, Inception1d, LeNet1d:**
  These models process raw PPG signals based one avergge of the targets (baseline) and using convolutional neural networks.
  ViewBaseLine, XResNet1d, Inception1d, & LeNet1d scripts
- **AlexNet1d & MiniRocket:**
  Additional raw time series models designed for fast feature extraction.
  View AlexNet1d & MiniRocket scripts
- **XResNet1d50+GNLL:**
  A variant of XResNet1d incorporating Gaussian Negative Log Likelihood Loss for enhanced BP regression performance.

D2 code repository: gitlab.com/qumphy/d2-code

README.md

A Systematic Evaluation of Uncertainty Quantification Techniques in Deep Learning: A Case Study in Photoplethysmography Signal Analysis

Paper here

This repository accompanies the paper *"A systematic evaluation of uncertainty quantification techniques in deep learning: a case study in photoplethysmography signal analysis"* by Ciaran Bench, Oskar Pfeffer, Vivek Desai, Mohammad Moulaeifard, Loïc Coquelin, Peter H. Charlton, Nils Strodthoff, Nando Hegemann, Philip J. Aston, and Andrew Thompson.

The work implements and evaluates **eight uncertainty quantification (UQ) techniques** on two clinically relevant PPG-based prediction tasks: **Atrial Fibrillation (AF) detection** (binary classification) and **blood pressure estimation** (regression of systolic and diastolic blood pressure, in both calibration-based and calibration-free settings).

Predictive Tasks

1. **Atrial Fibrillation (AF) Detection**
   - Binary classification: Predict presence of AF episode within a 25-second PPG segment.
2. **Blood Pressure (BP) Estimation**
   - Regression: Predict systolic (SBP) and diastolic (DBP) blood pressure from 10-second PPG segments.
   - Evaluated in two settings:
     - `calib`: Calibration data from same patients.
     - `calibfree`: No patient overlap (real-world generalization).

D4 code repository: gitlab.com/qumphy/d4-code

README.md

Machine-learning for Photoplethysmography Analysis:

Benchmark Problems and Benchmark Datasets 🔗

This repository contains the official code for the Benchmark Problems chosen and analysed during the QUMPHY project. The Benchmark Problems are presented in the D4 report of the project that is available online on the Qumphy homepage.

Benchmark Problem	Benchmark Dataset	Code
Blood pressure regression	VitalDB	D1-Code repository of Qumphy
	AuroraBP	this repository
Detection of Atrial Fibrillation	DeepBeat	D1-Code repository of Qumphy
	TriggersAF	this repository
	MIMIC-III-Ext-PPG	Physionet
	Liu2022	No extra scripts needed
Blood pressure classification	AuroraBP	this repository
Vascular age	AuroraBP	this repository
Detection of Sleep apnoea	OSASUD	this repository
	MESA	this repository
Repiratory Rate Regression	MIMIC-III-Ext-PPG	Physionet
	OSASUD	this repository
	MIMIC Perform Large	this repository

qumphy.ptb.de

gitlab.com/qumphy

qumphy-software.rtfd.io