Newsletter

March 18, 2026

What is QUMPHY About?

Need Photoplethysmography (PPG) signals are rich in information and easy to measure passively without any physical or mental limitations of the subject. As it is impossible for physicians to infer physiological parameters from the PPG signal by themselves, they need to rely on algorithms based on machine learning (ML) techniques for diagnosis. As of today, no regulations specifying how these ML algorithms have to be applied, how their performance has to be measured or how their associated uncertainties have to be specified exist.

Aim At the core of this project stands the development of measures to quantify the uncertainties associated with ML algorithms applied to medical problems, in particular the analysis and processing of PPG signals. To achieve this the following tasks will be addressed: (i) benchmark datasets will be generated using publicly available in vivo, and synthetic data (ii) different ML models and uncertainty quantification (UQ) methods will be used to analyse the processing of the PPG signals and specify the associated uncertainty and (iii) a good practice guide with accompanying software repository showcasing the used models, methods and benchmarks will be developed and made publicly available.

Objectives

The overall objective is to provide trustworthy machine learning models for analysing photoplethysmography signals in a medical context, by developing methods for the quantification of uncertainty in supervised machine learning and deep learning models applied to photoplethysmography signals and generating reference datasets to benchmark those models, supported by software being developed that will be publicly available for independent review of the models.

The specific objectives are:

1. To develop methods for quantifying the uncertainty for at least 3 existing classification and 3 existing regression supervised machine learning and/or deep learning models using photoplethysmography (PPG) data, considering the effects of both aleatoric (data) uncertainty and epistemic (model) uncertainty on model predictions.

2. To generate at least 5 measurement problems and their corresponding 5 datasets, using real and/or synthetic photoplethysmography data, that can be used to benchmark accuracy and uncertainty of supervised machine learning and deep learning models. In addition, to make those reference problems and datasets available to the medical device and digital health communities via an online repository.

3. To validate the uncertainties obtained for existing machine learning and deep learning models of Objective 1 and to compare the 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.

4. To engage with the medical device, digital and health communities to (a) promote the use of the good practice guide and the accompanying software repository through conference contributions, peer-reviewed journal articles and stakeholder workshops, (b) support the adoption of the benchmarking problems and datasets by providing guidelines for their use, and (c) develop a framework for independently reviewing machine learning models proposed by industry to assist them in getting regulatory approval.

5. To facilitate the take up of the technology and measurement infrastructure developed in the project by the measurement supply chain (NMIs, DIs, medical device calibration services), standards developing organisations (IEC, ISO), and end users (clinical practitioners, digital experts within the health communities, manufacturers of medical and healthcare products).

The Consortium

The consortium brings together the leading European metrology institutes (NMIs and DIs) in the fields of Machine Learning, Uncertainty Quantification and Medical Imaging, and they are complemented by a number of leading research institutes and companies that bring their specific knowledge and experience. In total, 6 NMIs/DIs, 8 universities or research institutes and 2 companies are involved in the project.

Setting the Benchmark for PPG & Machine Learning

How do we choose the best machine learning approach for heart health monitoring? As part of the QUMPHY project's mission to build trust in medical AI, our latest study in biomedical signal processing and control provides a much-needed roadmap. The team conducted a comprehensive "like-for-like" benchmarking of different input representations, comparing interpretable features, image-based scalograms, and raw waveforms. Testing these across two critical clinical use cases, blood pressure (BP) estimation and atrial fibrillation (AF) detection, the results are clear:

Raw Data Wins.

Deep neural networks operating on raw time series (signal-based) consistently outperform other representations.

Architecture Matters.
The strongest performance was observed in deeper convolutional neural networks (CNNs), such as XResNet.

Task-Specific Efficiency.
While high-capacity models led the way, the study reveals that smaller, lower-capacity CNNs remain highly competitive for specific clinical tasks.

This work serves as part of a good practice guide for developers and clinicians, ensuring that the next generation of wearable health technology is built on the most robust and accurate foundations possible.

Read the full paper here:
Machine-learning for photoplethysmography analysis: Benchmarking feature, image, and signal-based approaches, M. Moulaeifard, et al., Biomedical Signal Processing and Control, Volume 120 (2026) 10.1016/j.bspc.2026.109831.

How Much Can We Trust AI Predictions in PPG?

In clinical settings, an AI prediction is only as good as the confidence we can place in it. As part of our ongoing work in the QUMPHY project, we've released a new preprint on arXiv that systematically evaluates how well deep learning models "know what they don't know." The study provides an unprecedented evaluation of eight different uncertainty quantification (UQ) techniques applied to PPG signal analysis. Using the same critical use cases, AF detection and BP regression, we explored which methods best signal when a model's output might be unreliable. Key insights from the study are:

No "One Size Fits All".
The best UQ technique depends heavily on the specific clinical task and the metric being measured.

Beyond Global Metrics.
We found that assessing "local calibration" (how reliable a model is for a specific patient or measurement) provides much deeper insights than standard global averages.

Practical Roadmap.
The paper emphasizes that UQ evaluation must be tailored to the real-world use case, especially for wearables where we often have limited data per patient.

By quantifying uncertainty, we are one step closer to making AI-driven heart monitoring a reliable tool for doctors and patients alike.

Read the full preprint here:
A systematic evaluation of uncertainty quantification techniques in deep learning: a case study in photoplethysmography signal analysis, C. Bench, et al.
arxiv.org/abs/2511.00301.

3rd Stakeholder Workshop

On April 10, 2026, the QUMPHY consortium will host a dedicated stakeholder workshop focused on a single, vital goal: ensuring our research reaches the hands of those who need it most. After months of intensive development, we are thrilled to unveil and discuss the project's final results. Rather than offering a mere summary of our work, this session is designed to provide a clear, hands-on roadmap for how these outcomes can be utilised. We want to ensure that every participant obtains a deep understanding of the project's impact and the practical knowledge needed to integrate our findings into their daily workflows. By working together, we are moving closer to a future where heart health monitoring is not only smart but universally trusted.

Register for the workshop informally by sending a mail to nando.hegemann@ptb.de. We hope to see you there.

Publications

Peer Reviewed Articles

[1] A. D. Rodway, L. Hanna, J. Harris, R. Jarrett, C. Allan, F. Pazos Casal, B. C. Field, M. B. Whyte, N. Ntagiantas, I. Walton, A. Pankhania, S. S. Skene, G. D. Maytham, and 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". In: *eClinicalMedicine* 68 (02/2024), p. 102410. DOI: 10.1016/j.eclinm.2023.102410.

[2] M. Rinkevičius, J. Lázaro, E. Gil, P. Laguna, P. H. Charlton, R. Bailón, and V. Marozas. "Obstructive Sleep Apnea Characterization: A Multimodal Cross-Recurrence-Based Approach for Investigating Atrial Fibrillation". In: *IEEE Journal of Biomedical and Health Informatics* 28.10 (10/2024), pp. 6155–6167. DOI: 10.1109/jbhi.2024.3428845.

[3] L. Hanna, A. D. Rodway, P. Garcha, L. Maynard, J. Sivayogi, O. Schlager, J. Madaric, V. Boc, L. Busch, M. B. Whyte, S. S. Skene, J. Harris, and 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". In: *eClinicalMedicine* 75 (09/2024), p. 102788. DOI: 10.1016/j.eclinm.2024.102788.

[4] P. H. Charlton, V. Marozas, E. Mejía-Mejía, P. A. Kyriacou, and J. Mant. "Determinants of photoplethysmography signal quality at the wrist". In: *PLOS Digital Health* 4.6 (06/2025). Ed. by W. Karlen, e0000585. DOI: 10.1371/journal.pdig.0000585.

[5] C. Teichert, U. Hackstein, T. Krüger, and S. Bernhard. "Noninvasive detection of lower extremity artery disease using multi-site photoplethysmographic signals and machine learning". In: *npj Biosensing* 2.1 (07/2025). DOI: 10.1038/s44328-025-00044-z.

[6] M. Moulaeifard, P. H. Charlton, and N. Strodthoff. "Generalizable deep learning for photoplethysmography–based blood pressure estimation—A benchmarking study". In: *Machine Learning: Health* 1.1 (09/2025), p. 010501. DOI: 10.1088/3049-477x/ae01a8.

[7] C. Bench, V. Desai, M. Moulaeifard, N. Strodthoff, P. Aston, and A. Thompson. "Uncertainty quantification with approximate variational learning for wearable photoplethysmography prediction tasks". In: *Machine Learning: Health* 1.1 (10/2025), p. 015013. DOI: 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, P. J. Aston, P. H. Charlton, and N. Strodthoff. "Machine-learning for photoplethysmography analysis: Benchmarking feature, image, and signal-based approaches". In: *Biomedical Signal Processing and Control* 120 (07/2026), p. 109831. DOI: 10.1016/j.bspc.2026.109831.

Proceedings & Other

[9] P. Charlton and P. A. Kyriacou. "Wearable Photoplethysmography: Current Status and Future Challenges". In: *2023 Computing in Cardiology Conference (CinC)*. CinC2023. Computing in Cardiology, 11/2023. DOI: 10.22489/cinc.2023.076.

[10] N. Strodthoff. "Open Science to Foster Progress in Automatic ECG Analysis: Status and Future Directions". In: *2024 Computing in Cardiology Conference (CinC)*. Vol. 51. CinC2024. Computing in Cardiology, 12/2024. DOI: 10.22489/cinc.2024.057.

[11] M. Rinkevičius, P. H. Charlton, and V. Marozas. "Uncertainty in Photoplethysmography-Based Cuffless Blood Pressure Trend Monitoring: A Personalized Approach". In: *2024 Computing in Cardiology Conference (CinC)*. Vol. 51. CinC2024. Computing in Cardiology, 12/2024. DOI: 10.22489/cinc.2024.098.

[12] C. A. Bench, N. Strodthoff, M. Moulaeifard, P. Aston, and A. J. Thompson. "Towards Trustworthy Atrial Fibrillation Classification from Wearables Data: Quantifying Model Uncertainty". In: *2024 Computing in Cardiology Conference (CinC)*. Vol. 51. CinC2024. Computing in Cardiology, 12/2024. DOI: 10.22489/cinc.2024.068.

[13] P. Aston. "Does Skin Tone Affect Machine Learning Classification Accuracy Applied to Photoplethysmography Signals?" In: *2024 Computing in Cardiology Conference (CinC)*. Vol. 51. CinC2024. Computing in Cardiology, 12/2024. DOI: 10.22489/cinc.2024.038.

Preprints

[14] C. Bench, O. Pfeffer, V. Desai, M. Moulaeifard, L. Coquelin, P. H. Charlton, N. Strodthoff, N. Hegemann, P. J. Aston, and A. Thompson. *A systematic evaluation of uncertainty quantification techniques in deep learning: a case study in photoplethysmography signal analysis*. 2025. arXiv: 2511.00301 [cs.LG].

That's All

Thank you for reading. If you want to stay up-to-date with the project, you can always visit our project website:

www.qumphy.ptb.de

Acknowledgements

The project 22HLT01 QUMPHY has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

Funding for NPL, KCL, U Cambridge, U Surrey and Sector Health was provided by Innovate UK under the Horizon Europe Guarantee Extension.