Publishable Summary for 22HLT01 QUMPHY
Uncertainty quantification for machine learning models applied to photoplethysmography signals

Overview

Photoplethysmography (PPG) signals are rich in information and easy to measure passively without any physical or mental limitations of the subject. Machine learning (ML) algorithms are commonly used to extract physiological parameters from PPG signals for diagnosis, avoiding the need for complex and costly clinical review. 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. 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 are addressed: (i) benchmark datasets are generated using publicly available in vivo, and synthetic data (ii) different ML models and uncertainty quantification (UQ) methods are 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 is developed and will be made publicly available.

Need

A Photoplethysmograph measures the intensity of reflected or transmitted light through body tissue. Measured at almost any part of the human body (e.g., finger, wrist, arm, ankle, ear, neck), PPG signals change throughout the cardiac cycle as the changing volume of blood present in the tissue influences the intensity of the transmitted/reflected light. PPG signals are used both in a clinical setting and are collected by many wearable devices. They contain valuable information on the cardiovascular, respiratory, and autonomic nervous systems which is not yet routinely exploited. They are popular as they are easy to obtain non-invasively and PPG devices are cheap and widely available. Moreover, using smart devices or even contactless measurements by external cameras yields the possibility for long-term monitoring without restriction of patient comfortability. Until today, an algorithmic evaluation of PPG signals to infer physiological parameters or detect diseases is crucial for saving patients' lives but almost never used in clinical environments. One of the major reasons for this is the lack of trust in the output of any such algorithms.

Due to the vast amount of collected data, machine learning methodologies are essential for the extraction and evaluation of key features used for diagnosis. When applying machine learning in a medical context, however, confidence in the performance and predictions of the algorithms is particularly crucial since diagnostic mistakes can be fatal (false negative) or result in unnecessary anxiety and detrimental overtreatment (false positive). This is supported by the Executive Vice-President for a Europe fit for the Digital Age, Margrethe Vestager, who said "On Artificial Intelligence, trust is a must, not a nice to have". Hence an analysis of the uncertainty associated to ML algorithms and their predictions is indispensable to provide both clinicians and users of wearable devices with critical information about the quality and trustworthiness of the results produced. According to the US Center for Disease Control and Prevention (CDC) over 400 million people worldwide suffer from diabetes with predictions reaching 0.5 billion people worldwide by 2040. Similarly, the World Health Organisation (WHO) estimates that hypertension (elevated blood pressure) affects more than 1.2 billion adults between ages 30 and 79 worldwide with approximately half of the people being unaware that they have this condition or need treatment. A recent EC report states that: "By using digital solutions, such as wearables (…) citizens can actively engage in health promotion and self-management of chronic conditions. This in turn can help control the rising demand for health and care". Wearable devices such as smart watches allow early

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EURAMET. Neither the European Union nor the granting authority can be held responsible for them.

The project 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.

detection, monitoring and counteracting such conditions through lifestyle and diet changes or early medical treatment potentially saving millions of lives and billions in medical treatment costs. One of the major obstacles of this approach originates from the low signal quality obtained from PPG measurements through smart wearables at e.g., the upper side of the wrist. This again emphasises the need for both ML techniques to process the data, as well as reliable and trustworthy methods to quantify the uncertainties associated to the algorithms' predictions.

The goal of this project is to satisfy these needs by developing an environment, i.e., a good practice guide including a software framework for independent assessment of accuracy and uncertainty of ML algorithms (Objectives 1 and 3) and benchmark cases to test and compare ML algorithms against (Objective 2), to increase trust in ML applications for PPG signals and lay a foundation towards standardisation of ML in healthcare (Objective 4).

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).

Progress beyond the state of the art and results

Generate benchmark measurement problem datasets:

Accompanied by leading stakeholders the consortium identified relevant benchmark diagnostic tasks for PPG signals, such as blood pressure determination, detection of atrial fibrillation and classification of hypo- and hypertension. A list of accompanying datasets for the identified benchmark diagnostic tasks has been created and discussed with clinicians and medical device manufacturers. The datasets have been discussed in view of measurement quality, patient diversity and public accessibility.

The consortium evaluated the target communities' feedback and selected the 5 most promising out of the benchmark diagnostic tasks. The consortium formed benchmark specific subgroups and started to prepare the datasets for each of the benchmark problems. Once complete, the consortium will provide them publicly for the projects target user communities. This will be accompanied by publishing peer-reviewed research articles describing the chosen benchmarks and respectively selected datasets and provide examples on how to use them.

Develop methods for quantifying the uncertainty of supervised ML/DL models:

The consortium identified 2 prototypical benchmarks tasks, i.e., blood pressure determination and detection of atrial fibrillation, with accompanying datasets to develop and test machine learning and deep learning models. The models are developed for raw PPG signals, features extracted from the PPG signals and images created from PPG signals. A common evaluation framework to test and compare the performance of the developed methods is simultaneously developed by the consortium to guarantee good scientific practice. The consortium started to investigate several uncertainty quantification methods, such as Monte Carlo Dropout and Deep Ensembles, for their suitability to determine uncertainty for different machine learning algorithms with applications in PPG signals. The uncertainty quantification methods investigated include model dependent and independent approaches encompassing both aleatoric (data) and epistemic (model) uncertainties.

Once the machine learning models and uncertainty quantification methods have been implemented and tested, the consortium will test the performance and uncertainty budget for the selected benchmark diagnostic tasks. The tests will include an investigation of different demographic subgroups of the benchmark datasets, performance of the machine learning models on out-of-distribution data and training performance based on different quality of training data. The results of this evaluation will contribute to a good practice guide on uncertainty quantification for machine learning models applied to PPG signals.

Validation of uncertainties of ML/DL models for the benchmark problems:

The consortium has started to implement a common framework for performance evaluation of machine learning and deep learning algorithms. Additionally, uncertainty quantification methods have started to be developed, including model dependent and model independent methods and methods investigating both aleatoric (data) and epistemic (model) uncertainties. An online repository including the code base for the machine learning and deep learning models, the evaluation metrics and the uncertainty quantification methods has been created by the consortium to track the progress towards the projected outcome of the project. The repository adheres to common software development standards, includes automatic software testing and implements an automatically generated documentation of the repository code. The documentation is updated automatically if the code changes and openly available for public use.

After finishing the development of both machine learning and deep learning methods and uncertainty quantification methods, the joint code base will be used to validate and compare the performance of machine learning and deep learning models for the 5 chosen benchmark diagnostic tasks. The evaluation and comparison will be included in a good practice guide describing the techniques to quantify uncertainties for machine learning models applied to PPG signals. The respective evaluation scripts will be provided within the code repository and will be included as tutorials in the online software documentation.

Engage with medical device, digital and healthcare communities:

The consortium created a stakeholder committee at the start of the project which includes members from the medical device, digital and healthcare communities, which ensures that the benchmark scenarios considered are relevant for and needed by medical device manufacturers as well as clinicians. Additionally, the consortium held a stakeholder workshop to increase visibility of the project and to initiate discussions about future directions of the project. This first stakeholder workshop, conducted in month 3 of the project, provided an overview of the project aims and general structure of the work packages. The consortium held a second stakeholder workshop in month 10 of the project. In this workshop the invited stakeholders and other leading experts from clinics and digital healthcare companies discussed a set of diagnostic tasks involving PPG signals to use as potential benchmark problems. The workshop provided crucial input for the project and simultaneously increased its visibility.

At the end of the project, a good practice guide on uncertainty quantification for machine learning models using PPG signals will be created and distributed through the consortium and stakeholder committee to the projects target end user groups. The good practice guide will include a description of the selected benchmark problems and their respective datasets, a documentation and a user guide for the software repository containing the machine learning and deep learning models, uncertainty quantification methods and the common evaluation framework as well as tutorials showcasing good practice approaches to evaluate uncertainties for the implemented machine learning models and uncertainty quantification methods applied to the benchmarks defined in this project.

Outcomes and impact

Key dissemination and communication activities

In the first half of the project the consortium created a stakeholder committee and held two stakeholder workshops. The first stakeholder workshop was intended to foster dissemination of and communication about the project to the project's intended target communities. This was achieved by providing an overview of the project and its planned activities as well as detailed technical discussions about the next steps within the individual work packages. A second stakeholder workshop focussed on input from leading experts, such as clinicians or digital health professionals to ensure a reasonable and practically relevant setup of benchmark diagnostic tasks. Including stakeholders in the selection process improves one of the projects future key outputs and improves dissemination and outreach within the project's targeted end user groups. To improve the outreach of the project even further, the consortium created a website providing detailed information about the project and its projected outcomes. The website also contains download and publication areas providing official project communication documents, a project poster and flyer, the project's corporate design and information about the project's software repository, software documentation and other project related publications.

Outcomes for industrial and other user communities

PPG signals are collected by many wearable devices, such as smart watches, which are now widely available. By making digital health apps based on machine learning available on smart watches, individuals will be able to monitor different aspects of their own health. Combining uncertainty quantification with the machine learning predictions will ensure that only good quality predictions are presented to users which will in turn mean that they learn to trust the predictions and so will continue to use the app. In a hospital setting, pulse oximeters currently monitor patients' heart rate and blood oxygen continuously. By incorporating additional machine learning algorithms into these monitors, many more aspects of a patient's health and well-being could be monitored. Automatic detection systems that give unreliable alarms are often ignored, so the uncertainty quantification will be essential to provide some confidence in the alarms. This will provide continuous monitoring of all patients and will enable early detection of health deterioration.

Outcomes for the metrology and scientific communities

This project develops new methods for quantifying the uncertainty of machine learning predictions based on PPG signals. The benchmarking datasets will be of benefit to the metrological and scientific communities who may want to use these datasets with their own machine learning models or for other studies. Research papers will be submitted for publication in high impact peer-reviewed journals and the work in the project will be presented at relevant international conferences.

Outcomes for relevant standards

The consortium will contribute to national and international standards bodies and technical committees throughout the project, especially for AI in medicine. Special attention will be given to developing a good practise guide for uncertainty quantification of ML algorithms applied to PPG signals, which can act as a foundation for standardisation of PPG based medical applications in the future. The consortium anticipates high impact of the mathematical tools and advanced uncertainty quantification and propagation methods through international committees.

Longer-term economic, social and environmental impacts

Hypertension, diabetes, and myocardial infarction rank among the most common causes of death in human populations worldwide. Often, especially in the early stages of the diseases, a change in lifestyle and diet can be sufficient to mitigate these diseases, eradicating the need for expensive and possibly harmful treatment or medication. By making digital health apps based on machine learning available on smart watches, individuals will be able to monitor different aspects of their own health. Combining uncertainty quantification with the machine learning predictions will result in more trustworthy predictions. This project will contribute to the rapidly growing digital health industry which is providing wearable devices that enable individuals to continuously monitor their own health and well-being resulting in early detection of health issues or better management of chronic conditions. This will result in better health for users and, in some cases, to a longer life. As a wider impact, this project will provide a boost to Europe's rapidly growing digital health industry, leading to higher skilled employment and wealth for society, by providing digital healthcare companies as well as clinicians with an understandable and deployable good practice guide to assess the accuracy and uncertainty of ML algorithms in healthcare applications.

List of publications

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, 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, 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, https://doi.org/10.1016/j.eclinm.2024.102788

<table>
<tr><td colspan="2">Project start date and duration:</td><td>1 July 2023, 36 months</td></tr>
<tr><td colspan="3">Coordinator: Nando Hegemann, PTB Tel: +49 30 3481 7773 E-mail: nando.hegemann@ptb.de
Project website address: www.qumphy.ptb.de</td></tr>
<tr><td>Internal Beneficiaries:
1. PTB, Germany
2. CMI, Czechia
3. IMBiH, Bosnia and Herzegovina
4. IPQ, Portugal
5. LNE, France</td><td>External Beneficiaries:
6. FC, Belgium
7. FVB, Germany
8. KTU, Lithuania
9. THM, Germany
10. UGent, Belgium
11. Uni-Oldenburg, Germany</td><td>Unfunded Beneficiaries:
-</td></tr>
<tr><td colspan="3">Associated Partners:
12. KCL, United Kingdom
13. NPL, United Kingdom
14. SectorHealth, United Kingdom
15. SURREY, United Kingdom
16. UCAM, United Kingdom</td></tr>
</table>