

The QUMPHY Good Practice Guide

Philip Aston (NPL)

10/04/2026

Good Practice Guide

- The GPG describes methods for evaluating uncertainties for machine learning models applied to PPG signals
- It is intended to be an accessible and comprehensive guide to the work done in the project for those who are not technical specialists
- It is 69 pages long
- It was mostly written by Peter Harris (NPL)

Aim of the Good Practice Guide

- “This *Good Practice Guide* is aimed at *providers* of information derived from PPG signals, such as device manufacturers who want to ensure that model predictions are accompanied by reliable statements of uncertainty. It is also aimed at *users* of such information, such as clinicians who want to understand the basis of the provided information and how to interpret it. The audience of the GPG is not expected to be an expert in ML and UQ but to have a working knowledge of both topics and of some methods for ML and UQ.”

Good Practice Guide – Section 1

1	Introduction	1
1.1	Background	1
1.2	Aims and Organisation of the Good Practice Guide	3
1.3	Benchmark Problems.....	4
1.3.1	Determination of Systolic and Diastolic Blood Pressure.....	4
1.3.2	Detection of Atrial Fibrillation.....	5
1.3.3	Classification of Hypertension	5
1.3.4	Determination of Vascular Age.....	5
1.3.5	Detection of Sleep Apnoea.....	6
1.3.6	Determination of Respiratory Rate	6

Good Practice Guide – Section 2

2	Models for Supervised Learning	7
2.1	Background	7
2.2	Deep Learning Applied to Raw Signals.....	8
2.3	Deep Learning Applied to Image Representations	9
2.3.1	Continuous Wavelet Transform	10
2.3.2	Symmetric Projection Attractor Reconstruction Method	11
2.4	Machine Learning Applied to Extracted Features	13
2.4.1	Estimation of Blood Pressure	13
2.4.2	Detection of Atrial Fibrillation.....	16
2.4.3	Discrete Wavelet Transform	17
2.4.4	Signal Pre-Processing, Feature Selection and Ranking.....	18
2.4.5	Machine Learning Models	19

Good Practice Guide – Section 3

3	Uncertainty Quantification	20
3.1	Background	20
3.2	Model-Dependent Methods	21
3.2.1	Regression Tasks	21
3.2.2	Classification Tasks	24
3.3	Model-Independent Methods	27
3.3.1	Calibration Methods	27
3.3.2	Conformal Prediction	30

Good Practice Guide – Section 4

4	Validation of Uncertainty	33
4.1	Background	33
4.2	Evaluation Metrics for Classification Tasks.....	35
4.3	Evaluation Metrics for Regression Tasks.....	38

Good Practice Guide – Section 5

5	Benchmark Datasets	42
5.1	Determination of Systolic and Diastolic Blood Pressure	43
5.2	Detection of Atrial Fibrillation	43
5.3	Classification of Hypertension	45
5.4	Determination of Vascular Age	45
5.5	Detection of Sleep Apnoea	45
5.6	Determination of Respiratory Rate	46
5.7	Other Problems Studied	46
5.7.1	Classification of Skin Tone	46
5.7.2	Classification of Biological Sex	47

Good Practice Guide – Section 6

6	Software Repository	45
---	---------------------------	----

Good Practice Guide – Section 7

7	Ethics	47
7.1	Background	47
7.2	Ethics Reports.....	47
7.3	Ethics Self-Assessment	48
7.3.1	Involvement of Humans, Human Cells or Tissue	48
7.3.2	Personal Data.....	48
7.3.3	Animals.....	48
7.3.4	Non-EU Countries	48
7.3.5	Environment, Health and Safety.....	49
7.3.6	Artificial Intelligence	49
7.4	Potential Misuse of Results.....	51

Good Practice Guide – Section 8

8	Summary	51
8.1	Background	51
8.2	Benchmarking of Machine Learning Models	55
8.3	Benchmarking of Uncertainty Quantification Methods	56