

Non-invasive Photoplethysmography Blood Glucose Benchmark Dataset (PPGBG) v1.0 Release Agreement

Release Date: August 2025

Maintained By: School of Computer Science and Engineering, Changchun University of Technology

In Collaboration With: The First Hospital of Jilin University

Public Repository: <https://github.com/zacheryzhang/BG>

Corresponding Publication (Citation to be updated upon formal publication): Z. Zhang et al., “A Dynamic Incremental Learning Approach for Non-invasive Blood Glucose Measurement Using PPG Signals with an Open Benchmark Dataset,” 2025.

Introduction: The Non-invasive Photoplethysmography Blood Glucose Benchmark Dataset (PPGBG) is composed of fingertip photoplethysmography (PPG) signals, corresponding fingertip blood glucose measurements obtained via standard glucometers, and upper-arm cuff-based blood pressure (systolic/diastolic) readings. The data was collected from 183 subjects over a period of 285 days, featuring a blood glucose range of 4.2–16.3 mmol/L and a mean subject age of 34.17 ± 23.18 years. The dataset provides a dynamic feature landscape encompassing multi-session, multi-individual data that includes authentic physiological drift and external uncertainties.

Dataset Overview: The dataset contains raw PPG segments sampled at 1kHz, with a recommended segment length of 4 seconds, temporally aligned with their corresponding blood glucose values in mmol/L. It also includes synchronized blood pressure (SBP/DBP) and heart rate data. Furthermore, the dataset furnishes basic anonymized statistical characteristics such as height, weight, BMI range, and gender distribution, along with baseline health status and information on comorbidities, including cardiovascular and renal function.

Application and Access Protocol: Applications for access must be submitted using an official email address affiliated with the applicant's institution, such as a university, research institute, or medical facility. The submission must include a scanned PDF of this completed and signed agreement, accompanied by a brief research summary outlining the research objectives, methodology, anticipated outcomes, and the intended scope of data use; a detailed proposal is not required. All applications are subject to a case-by-case review. The maintenance team reserves the right to request supplementary materials or to deny an application without providing an explanation. Access will be granted via an authorized link, which remains valid for one month and must not be forwarded. Should new versions or patches of the dataset be released, the maintenance team may send update notifications to registered teams.

Scope of License: A non-exclusive, non-transferable, and revocable license is granted for research purposes only. This license is strictly limited to academic non-commercial research, including but not limited to algorithm evaluation, methodological validation, and inclusion in publications and dissertations. Use for educational demonstrations, competition baselines, or cross-modal transfer learning assessments is permissible only with separate prior written approval.

Prohibitions: Without explicit written permission, the following activities are strictly prohibited. The data shall not be used for commercial purposes, directly or indirectly, such as in commercial products or services, for training commercial models, or in paid consultations. Redistribution of the dataset, in whole or in part, by forwarding, sharing, or uploading to third-party platforms, cloud storage, or public code repositories is forbidden. Any attempt to reverse-identify subjects by linking the data with external information or employing methods like facial recognition, fingerprint analysis, or statistical attacks to reconstruct or infer personal identities is prohibited. The data shall not be used to train generative models (e.g., Diffusion, GANs, or LLM multimodal modules) that could produce samples potentially enabling the reverse-association of sensitive individual characteristics. Users must not attempt to recover expunged metadata, such as collection dates, specific locations, or equipment serial numbers. Furthermore, any activity that violates local medical, privacy, or data-transfer laws and regulations is forbidden.

Data Quality and Disclaimer of Warranty: The user acknowledges that the data may contain real-world noise, including physiological drift, motion artifacts, illumination disturbances, and measurement errors; researchers are responsible for assessing the data's suitability for their specific purposes. The maintenance team provides no express or implied warranties regarding the completeness, merchantability, or fitness for a particular purpose of the data. The user assumes all liability for any direct or indirect losses, legal risks, or ethical controversies that may arise from the use of the dataset. This data does not constitute medical advice or a clinical diagnosis, and any extension into clinical trials requires separate ethical board approval.

Publication and Citation Policy: Prior to any publication, the user must ensure that their work does not violate the confidentiality, anonymity, and compliance clauses of this agreement. For citation, the following format is temporarily recommended: Z. Zhang, H. Lu, B. Li, Y. Wang, J. M. Kurdestany, Q. Zhao, B. Liu, "A Dynamic Incremental Learning Approach for Non-invasive Blood glucose Measurement Using PPG Signals with an Open Benchmark Dataset," 2025.

Derivative and Reprocessed Data: Generated features and model parameters, excluding intermediate caches that could reversibly restore the original signals, may be publicly released with accompanying research papers, provided they do not contain any reversible mappings to the source data. If the original data is re-annotated (e.g., with event labels or task divisions), a statement assessing the potential risk of re-identification must be included upon its release. Large-scale hybrid datasets created by merging PPGBG with other datasets may not be independently re-published under the "PPGBG" name.

Revocation and Termination: The maintenance team may unilaterally terminate this license with written notice if any of the following occurs: discovery of unauthorized redistribution or suspected attempts at identity reconstruction; failure to provide proper citation as stipulated or refusal to rectify such omissions; or changes in the legal or policy environment that introduce risks associated with continued authorization. Upon termination, the user is required to destroy all copies of the data.

NAME OF INSTITUTION

PI SIGNATURE

DATE

ORGANIZATION ADDRESS

CONTACT INFORMATION

NAME OF RESEARCHER(s)

Please scan and email the completed agreement to luhuimin@ccut.edu.cn from your institutional email address. The subject line of the email should be "PPGBG Access Request - [Your Institution Name]". In the body of the email, please provide the brief research summary as described. If you do not receive a response within two weeks, you may send a single follow-up inquiry.