

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The Azumio Instant Heart Rate app is a commercial, freely-available smartphone application by Azumio Inc, which was used to collect PPG recordings. The PPG waveforms were pre-processed by an Azumio algorithm for camera artifact removal, utilizing standard de-trending and low pass filter techniques. Web-based data collection in the Health eHeart study is supported by custom code for the Health eHeart web portal and smartphone application.

Data analysis

The convolutional neural network was built in Python 2.7 using Keras (version 2.0.3) and TensorFlow (version 1.0.1). The logistic regression models and AUC were derived in SPSS v22.0 (IBM). Custom code for data processing and analysis (including the neural network) used in this study is copyright of the Regents of the University of California and can be made available through license.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study can be made available upon reasonable request from the authors, but restrictions apply to the availability of these data which were used under license for the current study and due to their containing information that could compromise participant privacy/consent.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For Primary and Contemporary cohorts, the largest sample size available during the inclusion period was used in order to maximize the sample size for neural network training and validation. Similarly to maximize validation sample size, we aimed for the largest enrollment in the Clinic Cohort given the time and resource constraints associated with prospective in-person enrollment.
Data exclusions	Participants in all cohorts were included if they made at least one PPG recording. Based on our prior data exploration and analysis using similar PPG data (doi:10.1038/s41746-019-0134-9), we prespecified exclusions for low sampling rate (<100 Hz; which is based on smartphone model), invalid date of birth or average PPG heart rate outside of 20-220 beats per minute which we assumed are not physiologic and thus predominantly erroneous/artifactual signals). PPG recordings of <5 seconds were excluded, as were individuals who answered "I don't know" to the self-reported diabetes question.
Replication	We replicated our results in a total of three test datasets, including a hold-out test dataset from the Primary cohort, a temporally distinct cohort enrolled into Health eHeart after initial data-lock ("Contemporary Cohort"), and a prospectively enrolled in-person clinic cohort ("Clinic Cohort").
Randomization	Participants in the Primary cohort were divided into training, development and test datasets randomly, and all datasets are disjoint.
Blinding	Investigators were not blinded to group allocation, however all three test datasets were kept separate and not used during model development. Blinding is not relevant in this study since the exposure of deployment of the trained neural-network on test dataset(s) PPG data occurs without investigator input.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	A total of 3,564 participants (6.6%) had self-reported diabetes and 50,306 (93.4%) did not have self-reported diabetes. Compared to those without diabetes, those with diabetes were older (mean±SD): 54.6±14.7 vs 45.0±15.1, p<0.001), more likely male (59.3% vs 52.8%; p<0.001), had a higher BMI (32.1±7.0 vs 27.3±5.9; p<0.001), less likely non-Hispanic whites (Table 1) and had higher HR (83.8±14.5 vs 79.9±15.1 bpm; p<0.001).
Recruitment	Health eHeart (HeH) Study is a worldwide, internet-based, longitudinal eCohort; English-speaking adults, 18 years or older, with an email address were eligible to join. Participants were actively recruited through a variety of campaigns at UCSF (through clinics and electronically delivered invitations) and by partner organizations (e.g., American Heart Association), and passively recruited through word of mouth and press releases. Existing users of the Azumio Instant Heart Rate app were also invited to join Health eHeart. Since all study participants voluntarily enrolled into Health eHeart and elected to download the Azumio smartphone app, there is likely to be selection bias in our Primary and Contemporary cohorts, as we describe in limitations. The Clinic Cohort enrolled consecutive patients referred to three cardiovascular prevention clinics who consented to participate in the study. We opted to perform our additional validation in the Clinic Cohort precisely because it does not likely

exhibit the selection bias of the model derivation cohort, providing a test of external generalizability of the neural network performance.

Ethics oversight

The University of California San Francisco Institutional Review Board approved this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.