nature portfolio

Corresponding author(s): Honghua Yu, Wing W.Y.Ng, Ching-Yu Cheng, Zhuoting Zhu

Last updated by author(s): Mar 12, 2026

# Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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<br>*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<br>*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**

For the model training, we used fundus images from UK Biobank and three hospital-based centers in China to maximize ethnic and demographic heterogeneity in the training dataset. For external test, we utilized datasets from distinct healthcare settings in China and a multi-ethnic cohort from Singapore. The resource-limited dataset was retrospectively assembled from three clinical sites in China: the First Affiliated Hospital of Xinjiang Medical University, Linzhi People's Hospital of Tibet, and the People's Hospital of Guangxi Zhuang Autonomous Region. Additionally, we evaluated the model on the Singapore Epidemiology of Eye Diseases (SEED) study, a population-based cohort encompassing Chinese, Malay, and Indian adults in Singapore. The prospective silent trial was conducted between September and October 2025 to evaluate the technical robustness and seamless integration of the Reti-Pioneer framework in Guangdong, China.

**Data analysis**

We developed a multimodal learning model that integrates fund us images of varying image quality with structured clinical metadata. This architecture ensembles large-scale vision foundation models, including Swin Transformer, Vision Mamba, and RETFound, to leverage their complementary capabilities. The code being used in the current study for developing the algorithm is provided via GitHub at https://github.com/lyhyl/Reti-Pioneer. Discriminative performance is evaluated using area under the receiver operating characteristic curve (AUROC) Comparisons between AUROCs were conducted using Delong's test. Calibration is assessed using calibration plots (observed vs. predicted risk by deciles). Overall performance is evaluated using Brier score. Sensitivity, specificity, positive predictive value and negative predictive value for classification are reported with 95% confidence intervals. Confidence intervals were obtained using the percentile bootstrap method on 1000 bootstrap samples. To evaluate clinical utility across a range of risk thresholds, decision curve analysis was considered.
Data analyses and visualizations were conducted under Python (v.3.12) and R (v.4.2.0). Reti-Pioneer was implemented on the PyTorch (version 2.8.0) and optimized by AdamW optimizer with a warm-up strategy and a cosine annealing learning rate scheduler. The construction, fine-tuning, and testing processes were performed on an RTX 3080 GPU (10GB dedicated memory, CUDA version 12.8).

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Individual-level patient data can be accessible with the informed consent of the Data Management Committee from institutions and are not publicly available. Requests for access to deidentified individual-level data from the China can be submitted via email to Prof Honghua Yu (yuhonghua@stu.edu.cn) with detailed proposals for approval. SEED data might be made available to researchers who meet the criteria for access to confidential data and upon Institutional Review Board's approval, requests for access can be made to Prof Ching-Yu Cheng (chingyu.cheng@nus.edu.sg). Investigators who consent to the terms of the data transfer agreement, including, but not limited to, the use of these data only for academic purposes, and to protect the confidentiality of the data and limit the possibility of identification of participants, will be granted access. The data from the UK Biobank can be obtained via controlled access on their web portal (https://www.ukbiobank.ac.uk/).

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | Throughout this study, the variable sex was employed to denote biological attributes (categorized as female or male based on standard clinical documentation), which were essential for baseline participant characterization. The proportions of female and male participants are detailed in the primary demographic summary (Table 1, Supplementary Table 1 & 8). |
| Reporting on race, ethnicity, or other socially relevant groupings | All participant demographic data were collected via self-reported ethnicity questionnaires administered at the time of enrollment. No other socially relevant or intersectional groupings, such as socioeconomic status, health literacy level, primary language were systematically recorded or utilized for subgroup analysis in this study. |
| Population characteristics | The population characteristics is described in Supplementary Table 1 & 8, and Table 1. |
| Recruitment | The prospective trial was conducted between September and October 2025 to evaluate the technical robustness and seamless integration of the Reti-Pioneer framework within real-world clinical workflows under blinded conditions. To this end, eligible participants were systematically recruited to simulate a realistic clinical environment. The enrolled participants, while clinically indicated for screening, might not fully represent the entire spectrum of disease severity, particularly very early or advanced pathologies rarely seen in routine screenings. |
| Ethics oversight | This study was conducted with approval from the Institutional Review Board of Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences (KY-N-2022-134). For the UK Biobank data component, ethical approval was obtained through the NHS National Research Ethics Service (Ref: 11/NW/0382; Project ID: 86091). All participants provided written informed consent prior to study participation. To ensure confidentiality, all data were systematically de-identified by removing all personal identifiers prior to analysis. |

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

# 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

# Life sciences study design

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

| | |
|---|---|
| Sample size | The framework was tested on an internal dataset comprising 107,730 images from 53,865 participants, and further validated on independent datasets across resource-limited and high-resource regions in China, as well as muti-ethnic cohort in Singapore, collectively comprising 23,232 images from 11,616 participants. We further evaluated the Reti-Pioneer framework on a fully withheld longitudinal subset from the UK Biobank(Jan 2006-Dec 2021) to assess its performance in predicting six endocrine diseases over 5- and 10-year intervals. The analysis included 15,704 participants (31,748 images). |
| Data exclusions | For the prediction tasks, individuals with prevalent disease at baseline were excluded. |
| Replication | For external test, we utilized datasets from distinct healthcare settings in China and the multi-ethnic cohort Singapore Epidemiology of Eye Diseases study (SEED). The resource-limited dataset was retrospectively assembled from three clinical sites in China: The First Affiliated Hospital of Xinjiang Medical University, Linzhi People's Hospital of Tibet, and the People's Hospital of Guangxi Zhuang Autonomous Region. |

This dataset comprised 1,373 participants (2,746 images) with dilated pupils. The high- resource dataset comprised retrospectively collected data from physical examination centers (2,003 participants; 4,006 images) and large tertiary hospitals (587 participants; 1,174 images), captured predominantly under non-dilated conditions. Additionally, we evaluated the model on the SEED study, a population-based cohort encompassing Chinese, Malay, and Indian adults in Singapore, as previously described[27]. A total of 15,306 CFPs from 7,653 Singaporean adults (2,800 Chinese, 2,385 Malay, and 2,468 Indian) were included in the external test set.

Randomization: NA

Blinding: Participated patients and consultant-level reviewer will be masked.

# 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 |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☐ | ☒ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration: NA

Study protocol: NA

Data collection: NA

Outcomes: NA

## Plants

Seed stocks: NA

Novel plant genotypes: NA

Authentication: NA