

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

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | 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 code used to train, fine-tune and evaluate RETFound from Yukun Zhou is available at https://github.com/rmaphoh/RETFound_MAE , which is based on PyTorch. All pre-trained model weights are available at https://huggingface.co/YukunZhou . Images were processed with automated retinal image analysis tool AutoMorph v.1.0 (https://github.com/rmaphoh/AutoMorph).
Data analysis	Data was analysed with Python v3.11.0 (https://www.python.org/), NumPy v1.26.4 (https://github.com/numpy/numpy), SciPy v1.15.2 (https://www.scipy.org/), Matplotlib v3.8.4 (https://github.com/matplotlib/matplotlib), pandas v1.5.0 (https://github.com/pandas-dev/pandas), Scikit-Learn v1.4.2 (https://scikit-learn.org/stable), Pillow v10.2.0 (https://pypi.org/project/Pillow).

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](#)

The MEH-AlzEye data consists of routinely collected healthcare data. Owing to their sensitive nature, the dataset is subject to controlled access by means of a

structured application process. The AlzEye dataset is subject to the contractual restrictions of the data sharing agreements between National Health Service Digital, Moorfields Eye Hospital and University College London, and is not available for access beyond the AlzEye research team. National and international collaborations are welcomed, although restrictions on access to the cohort mean that only the AlzEye researchers can directly analyse individual-level systemic health data. Interested collaborators should contact the chief investigator P.A.K.

For Shanghai-SDPP data, individual-level patient data can be accessible with the consent of the data management committee from institutions and are not publicly available. Requests for the non-profit use of the fundus images and related clinical information should be sent to T.Y.W. The data management committee will then review all the requests and grants (if successful). A formal data transfer agreement will be required upon approval. Generally, all these requests for access to the data will be responded to within 1 month.

Data for ocular disease experiments are publicly available online and can be accessed through the following links: IDRID (<https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid>), MESSIDOR2 (<https://www.adcis.net/en/third-party/messidor2/>), APTOS2019 (<https://www.kaggle.com/competitions/aptos2019-blindness-detection/data>), Glaucoma Fundus (<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1YRRAC>), JSIEC (<https://zenodo.org/record/3477553>), and Retina (<https://www.kaggle.com/datasets/jr2ngb/cataractdataset>).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research.](#)

Reporting on sex and gender

Biological sex information for MEH-AlzEye and Shanghai-SDPP was collected via self-report. MEH-AlzEye includes 353,157 patients (190,494 female and 162,663 male). Shanghai-SDPP includes 79,284 participants. Experiments were conducted both on female and male. We used part of MEH-AlzEye and Shanghai-SDPP data to develop parallel foundation models, respectively, and held-out data for downstream validation (detailed in Supplementary Table 2).

Population characteristics

MEH-AlzEye is a retrospective cohort study linking ophthalmic data of 353,157 patients, who attended Moorfields Eye Hospital between 2008 and 2018, with systemic health data from hospital admissions across the whole of England. Systemic health data are derived from Hospital Episode Statistics (HES) data relating to admitted patient care (APC), with a focus on cardiovascular disease and all-cause dementia. More details can be found in the method section. SDPP cohort was drawn from a community-based longitudinal study of 79,284 participants who underwent physical examinations between December 2015 and November 2022.

Recruitment

MEH-AlzEye is a retrospective cohort study linking ophthalmic data of 353,157 patients who attended Moorfields Eye Hospital between 2008 and 2018. Shanghai-SDPP is a retrospective cohort study linking ophthalmic data of 79,284 participants who underwent physical examinations at Huadong Sanatorium and Shanghai Sixth People's Hospital between December 2015 and November 2022.

Ethics oversight

This study involves human participants and was approved by the London-Central Research Ethics Committee (18/LO/1163, approved 1 August 2018), Advanced statistical modelling of multimodal data of genetic and acquired retinal diseases (20/HRA/2158, approved 5 May 2020), Confidential Advisory Group for Section 251 support (18/CAG/0111, approved 13 September 2018), and the Ethics Committee of Shanghai Sixth People's Hospital (Approved No: 2019-087, approved 29 August 2019). The National Health Service Health Research Authority gave final approval on 13 September 2018. Moorfields Eye Hospital NHS Foundation Trust validated the de-identifications for MEH data. Only de-identified retrospective data were used for research.

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

Data for developing parallel foundation models was sampled from MEH-AlzEye and Shanghai-SDPP (each contributing 904,170 CFPs). Data for ocular disease diagnosis were from public datasets, detailed in Supplementary Table 1. Data for comparing intra-site and inter-site adaptation were from held-out MEH-AlzEye and held-out Shanghai-SDPP, introduced in Supplementary Table 2. Datasets were chosen based on the availability of labels that would permit model comparison, which is dependent on the specific clinical task being evaluated. The chosen validation datasets were deemed to be suitable based on their parameters, which are summarised Supplementary Table 1 Dataset characteristics. Formal sample size calculations were not performed due to the lack of established methods when applied to machine-learning classification studies.

Data exclusions

Data failed image processing with AutoMorph were excluded. Data without systemic health labels were excluded. For more details please refer to the method section.

Replication	All patients were randomly selected and were not correlated in any way. The replication of experiment results were confirmed in 5 times with 5 different random seeds.
Randomization	The training/validation/testing data for downstream tasks were randomly splitted in ratio of 55%:15%:30%. For each patient, we only included one eye data from one visit to avoid potential bias by inconsistent individual visits.
Blinding	When assigning patients randomly to training, validation and testing groups investigators were blinded to patient covariates and all features in the dataset not required to perform the research.

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 and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

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