nature portfolio

Corresponding author(s): Nikolas Pontikos

Last updated by author(s): Sep 29, 2022

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 *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	All the software relating to this project is under the Github organisation: https://github.com/Eye2Gene/. The code for training/testing the classification model is in the private Github repository: https://github.com/Eye2Gene/ClassificationModel . The code for the web app is in the private Github repository: https://github.com/Eye2Gene/eye2gene_app . The running version of the web app is accessible at https://app.eye2gene.com and the reviewers are able to login with the following credentials: user: reviewer@nature.com password: naturemedicine1995 All code can be made available upon request to the editor or the reviewers.
Data analysis	Same as above.

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 data is stored at https://grading.readingcentre.org and can be made available for viewing at the request of the editors/reviewers.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Gender information has been collected for the participants as recorded in their clinical record.
Population characteristics	ge at presentation, self-reported ethnicity, gender, mode of inheritance of condition and diagnostic gene associated with the condition.
Recruitment	Retrospective deidentified data collected during standard clinical care.
Ethics oversight	UK Health Research Authority - IRAS 242050 - REC 22/WA/0049 https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/eye2gene-10/

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	Since we are dealing with a rare condition, the entirety of the IRD data available at Moorfields Eye Hospital was use for the training.
Data exclusions	Participants were excluded based on availability/quality of imaging or missing genetic diagnosis information. The participant list was further filtered to only include participants with a gene diagnosis in one of 36 genes.
Replication	External validation was conducted in 4 external datasets.
Randomization	During the Eye2Gene AI training process participants were randomized to 5 different folds.
Blinding	Human benchmarking and evaluation of the Eye2Gene AI algorithm involved blinding the expert ophthalmologists and the AI algorithm to the known diagnostic gene when presented with a retinal scan.

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

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	Z6364106/2021/11/67, Data Protection reference number
Study protocol	The study protocol is available for download here: https://liveuclac-my.sharepoint.com/:w:/g/personal/rmhanpo_ucl_ac_uk/EZmNh0tHYHtBgWfHr-Tj758BW0R7sb1t7HvOI19kwT8T4A?e=EhzFaE
Data collection	Retrospective de-identified data collection over 2006-06-05 to 2018-04-05.
Outcomes	The measured outcomes are the Eye2Gene AI model prediction accuracy, top-5 accuracy and area under the receiver operating characteristic curve for identifying the correct diagnostic gene based on retinal scans.