nature portfolio

Corresponding author(s): Clarisse de Vries

Last updated by author(s): Mar 18, 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	Mia™ version 2.0.1 (proprietary) was used to generate the recall decision. Proprietary Python code was used to optimise the Mia™ decision threshold. Custom R code was used to generate the evaluated dataset.
Data analysis	Custom R code was used to analyse the data in this study.

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

Access to the raw SBSS data and mammograms (de-identified participant data) is subject to the required approvals (e.g. PBPP, NHS R&D, REC approval) and data agreements being in place. More information can be found on the DaSH website: https://www.abdn.ac.uk/iahs/facilities/grampian-data-safe-haven.php

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	Minimum sample size was calculated by an independent statistician commissioned (Quantics, CRO: Veristat). Sensitivity, specificity and recall rate with corresponding 95% confidence intervals (exact Clopper-Pearson) were planned. The aimed precision of the 95% confidence interval is 5% for both sensitivity and specificity and 1% for recall rate. Assuming sensitivity and specificity at 80%, 264 confirmed positive cases and 264 confirmed negative cases were required for each analysis. For an estimated 15% recall rate, 85% power was estimated with 5,150 cases. No sample sizes were calculated for exploratory endpoints.
Data exclusions	The initial extracted dataset was filtered to remove screenings without mammographic images (due to failed anonymisation), self-referrals, and cases from women outside the UK breast screening age range (50 to 70 years). The dataset for evaluation was limited to a complete 3-year UK screening cycle (1st of April 2016 to 31st of March 2019) prior to the application of recommended exclusions6: screenings with a "technical recall" reader opinion due to an image quality issue with one or more of the mammograms, women who had a previous mastectomy, women with complex physical requirements which could limit image quality, and cases which do not have all four standard image views. One recommended exclusion, prior breast surgery, was not incorporated as the measure available in the dataset was self reported and therefore difficult to interpret consistently.
Replication	In this study, we attempted to replicate previously reported performance of an AI algorithm with data from a different clinical site (not used for training the algorithm), to establish its generalisability. Our results show that the AI performance did not transfer without local optimisation of the decision threshold.
Randomization	Participants were not allocated to experimental groups.
Blinding	The AI vendor, who produced the AI predictions, was blinded to the cancer outcomes. The subsequent analysis of the AI performance was performed independently of the vendor.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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