nature portfolio

Corresponding author(s): Shekoofeh Azizi, Alan Karthikesalingam and Vivek Natarajan

Last updated by author(s): Apr 27, 2023

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	No software was used for data collection.
Data analysis	Several major components of the work are available in open-source repositories, such as the Tensorflow library (https://www.tensorflow.org). The code-base and pretrained weights used for self-supervised pre-training are available at SimCLR GitHub (https://github.com/google-research/simclr). The code-base and pretrained weights for the BiT models are available at Big Transfer GitHub (https://github.com/google-research/big_transfer). All experiments and implementation details are described in sufficient detail in Methods and in the Supplementary Information to support replication with non-proprietary libraries. The codebase used for our comparison to ResNet-RS was based on the ResNet-RS GitHub repository (https://github.com/tensorflow/tpu/tree/master/models/official/resnet/resnet_rs). A number of the checkpoints and models generated through REMEDIS are readily accessible to researchers via the PhysioNet open access project (https://physionet.org/projects/N0nzs56nBt1IBM073DOr). Additionally, the Foundation Medical ML repositories on GitHub offer access to codes that can be used to train REMEDIS-based models.

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 datasets from Northwestern Medicine and Apollo Hospitals were used under a licence for the current study, and are not publicly available. Applications for access to the Optimam database can be made using this web form (https://medphys.royalsurrey.nhs.uk/omidb/getting-access). The de-identified tele-dermatology data used in this study are not publicly available owing to restrictions in the data-sharing agreement. The unlabelled dataset used for DME classification is de-identified data from EyePACS Inc. Interested researchers should contact jcuadros@eyepacs.com to enquire about access to EyePACSdata and approach the Office of Research and Development (https://www.research.va.gov/resources/ORD_Admin/ord_contacts.cfm) to enquire about access to VA data. The rest of annotated data for ID and OOD DME classification tasks are collected at the Rajavithi Hospital Thailand and at the Lions Eye Institute and are not publicly available owing to restrictions in the data-sharing agreement. Data used in the evaluation and pre-training of the chest-X-ray-condition classification, including MIMIC-CXR (https://physionet.org/content/mimic-cxr/2.0.0), CheXpert (https://stanfordmlgroup.github.io/competitions/chexpert), and ChestX-ray14 (https://www.kaggle.com/datasets/nih-chest-xrays/data), are publicly available. Data used for the ID fine-tuning and evaluation of the detection of metastases are publicly available on the CAMELYON challenge website (https://camelyon16.grand-challenge.org/Data). The Cancer Genome Atlas (TCGA) data has been used for pre-training for both the pathology-based metastases-detection and survival-prediction tasks, and are available via the NIH website (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). The rest of the data used in pathology tasks are not publicly available, owing to restrictions in the data-sharing agreement. Moreover, ImageNet-1K (ILSVRC2012) [68] has been used for the pre-training of baseline supervised models, and ImageNet-21K has been used for the pretraining of BiT-M models. Both of these are publicly available via the ImageNet website (https://image-net.org/download.php). Please note that ID and OOD below refer to "in distribution" and "out of distribution", respectively

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	Sex subgroup reporting is in Supplementary Fig. 8. 'Sex' (rather than gender) is the right term because the datasets contain information on male/female as categories.
Reporting on race, ethnicity, or other socially relevant groupings	Age and breast-density subgroup reporting is provided in Supplementary Fig. 8. Individual race/ethnicity data points were not consistently available.
Population characteristics	Age, sex and other relevant population information are available in these papers: Liu, Y. et al. A deep learning system for differential diagnosis of skin diseases, Nature medicine, 26, 900–908 (2020). McKinney, S.M. et al. International evaluation of an AI system for breast cancer screening, Nature, 577, 89–94 (2020).
Recruitment	Not applicable, as we used retrospective de-identified data.
Ethics oversight	The institutional-review-board waiver for this study on retrospective de-identified data was obtained from Advarra IRB.

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	Test sets for all 7 medical tasks have been previously published, and the sample size has therefore been shown to be sufficient for the estimation of a model's diagnostic accuracy with acceptable uncertainty.
Data exclusions	The datasets used reflect those used in previous papers; no additional data exclusions were applied.
Replication	The code of is available to ensure that the work can be replicated externally. Internal tests for replication are a standard engineering practice at Google.
Randomization	Randomization was performed to create the training, test and validation splits of the different datasets used in the study.

Blinding	Blinding was not applicable to the study.

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