nature portfolio

Corresponding author(s): Yang Yang, PhD

Last updated by author(s): Jun 7, 2021

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	GE Centricity PACS-IW 3.7.3.9: FDA-approved fully features PACS viewer. Used to collect reader study results and key images
Data analysis	Pandas (version 0.23.4): open source library Pandas was used for tabular data Numpy (version 1.20): open source library Numpy was used for tabulated data Matplotlib (version 2.2.2): open source library Matplotlib was used for making plots sklearn (version 0.20.2): open source library scikit-learn was used for metrics such as AUC Tensorflow (version 2.4.0): open source library Tensorflow was used to train deep learning models Statsmodels (version 0.10.2): a python library Statsmodels was used to calculate the confidence intervals, normality of distribution, and p-values for mean absolute error pROC (version 1.16): libray pROC in R was used to calculate the confidence intervals for AUC and calculate the p-values of AUC

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 RadImageNet dataset was licensed from RadImageNet LLC (formed by East River Medical Imaging, PC) to Icahn School of Medicine at Mount Sinai for this research.
The five medical applications datasets are available at:
Bone age prediction: https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pediatric-bone-age-challenge-2017
Pneumonia detection:
https://www.kaggle.com/c/rsna-pneumonia-detection-challenge
ACL tear detection:
https://stanfordmlgroup.github.io/competitions/mrnet/
SARS-CoV-2 detection:
http://ncov-ai.big.ac.cn/download?lang=en
Intracranial hemorrhage detection:
https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection

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	We collected a total of 203,341 CT (52,691), MRI (142,422), and ultrasound (8,228) studies from 131,872 patients between 1 January 2005 and 31 January 2020 where patients had diagnostic scans at East River Medical Imaging in New York. Over 1.4 million images that consisted of 356,525 CT images, 672,676 MRI images, and 389,885 ultrasound images, were included in this study. A total of 3 modalities, 14 anatomies and 165 labels were used for model development. We split the dataset into 75% training set (1,129,987 images), 10% validation set (118,501 images) and 15% test set (170,596 images) stratified by patient ID to train convolutional neural networks from scratch. The model evaluation was performed on the test set. All the images in the dataset were independent and non-repeating samples. The test set was a hold-out set that was unseen by the models. We believe the sample size was sufficient for the test set because the test set represent a large sample size and was randomly sampled. For the other five transfer learning application, a total of 1,021 images were used for MRI ACL tear detection, a total of 12,611 images were used for x-ray bone age prediction, and a total of 26,684 chest x-rays for pneumonia detection, a total of 58,766 images for chest CT SARS-CoV-2 detection, and 573,614 images for CT head hemorrhage detection.
Data exclusions	The data exclusion criteria was pre-established. We excluded data when images have low equality (motions and/or artifacts). A total of 4 MRI studies were excluded because of motion for the ACL tear study.
Replication	We created a hold-out test for each deep learning model, including the models derived from scratch and the 4 transfer learning application models except ACL tear detection. A 5-fold cross validation was conducted for MRI ACL tear detection due to small sample size.
Randomization	We stratified the dataset by patient ID or study ID and randomly split the dataset into 75% training set, 10% validation set, and 15% test set for models derived from scratch and for the five transfer learning applications. All the cases in the dataset were independent and non-repeating samples.
Blinding	All experiments are based on de-identified studies. Blinding is not necessary.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	We collected a total of 203,341 CT (52,691), MRI (142,422), and ultrasound (8,228) studies from 131,872 patients between 1 January 2005 and 31 January 2020 where patients had diagnostic scans at East River Medical Imaging in New York.
Recruitment	No patients recruitment was performed. All studies performed at East River Medical Imaging that were annotated by fellowship trained radiologists for the pre-established collecting period were collected and analyzed.
Ethics oversight	East River Medical Imaging, PC and Icahn School of Medicine at Mount Sinai

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