

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 type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" 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 type="checkbox"/> | <input checked="" 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

Data analysis

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 two chest x-ray datasets used in this study are publicly available. The CheXpert dataset can be obtained after signing a data use agreement by following the instructions at <https://stanfordmlgroup.github.io/competitions/chexpert/>. The MIMIC-CXR dataset can be obtained after signing a data use agreement and

completing a credentialing process by following the instructions at <https://physionet.org/content/mimic-cxr/2.0.0/>. For the MIMIC-CXR dataset, the patient race information can be obtained via the admissions table in the MIMIC-IV dataset: <https://physionet.org/content/mimiciv/2.2/>.

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	We use two public chest x-ray datasets in this work: CheXpert and MIMIC-CXR. CheXpert consists of 45% Female patients and 59% Male patients. MIMIC-CXR consists of 49% Female patients and 51% Male patients.
Reporting on race, ethnicity, or other socially relevant groupings	Race was defined based on self-report. CheXpert consists of 64% white patients, 12% Asian patients, 5% Black patients, and 19% Other or Unknown. MIMIC-CXR consists of 67% white patients, 17% Black patients, 4% Asian Patients, and 12% Other or Unknown.
Population characteristics	The mean age of patients in CheXpert is 60.3 years old (s.d. of 18.6 years). The mean age of patients in MIMIC-CXR is 63.0 years old (s.d. of 17.7 years).
Recruitment	The CheXpert dataset was constructed based on chest x-ray exams that were performed at the Stanford University Medical Center between 2002-2017 in both inpatient and outpatient settings. The MIMIC-CXR dataset was collected from the Beth Israel Deaconess Medical Center (BIDMC) and was constructed by first querying the BIDMC electronic health record (EHR) to obtain a list of patients who received a chest radiograph in the emergency department from 2011-2016. All chest radiographs available in the BIDMC Radiology Information System (RIS) for this set of patients from 2011-2016 were then retrieved.
Ethics oversight	All analyses were performed on de-identified, public datasets. The CheXpert dataset can be obtained after signing a data use agreement by following the instructions at https://stanfordmlgroup.github.io/competitions/chexpert/ . The MIMIC-CXR dataset can be obtained after signing a data use agreement and completing a credentialing process by following the instructions at https://physionet.org/content/mimic-cxr/2.0.0/ . The study is classified as not-human subjects research as determined by the Dana-Farber/Harvard Cancer Center Institutional Review Board.

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](https://www.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	Two large public datasets were used for all analysis: CheXpert which consists of 224,316 chest x-ray images from 65,240 patients and MIMIC-CXR (MXR) which consists of 377,110 chest x-rays from 65,379 patients.
Data exclusions	In analysis by patient race, we consider subgroups of Asian, Black, and white patients because there are insufficient numbers of images from patients of other self-reported races to effectively analyze. In the MIMIC-CXR dataset, there were a small proportion of files (0.3%) which were excluded because there were corrupted and could not be read.
Replication	Confounder analysis was performed based on resampling approaches and DICOM-based evaluation, where similar results were observed across all analyses.
Randomization	Each dataset was split randomly at the patient level into training, validation, and testing splits with percentages of 70/10/20% respectively.
Blinding	The AI models were only trained on data from the training set. All results were reported on the testing set.

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 | Involvement 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 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

Methods

- | n/a | Involvement 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 |