nature portfolio

Corresponding author(s): David Chen, Deborah Kwon

Last updated by author(s): Mar 30, 2026

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	pydicom 2.20, pandas 2.0, nltk 3.8.1
Data analysis	python 3.9, pytorch 2.1, transformers (from huggingface), https://github.com/Makiya11/CMRCLIP

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 supporting the findings from this study are available within the manuscript and its supplementary information. The ACDC dataset we used for external validation is publicly accessible at (https://www.creatis.insa-lyon.fr/Challenge/acdc). The raw data from Cleveland Clinic are not publicly available due to inclusion of protected health information. Those interested in collaborative research may contact the corresponding authors to establish a data use agreement to facilitate limited data sharing.

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 report split of sex in our datasets. No separate sex or gender analysis were performed as subpopulation bias was not in the scope of this work.
Reporting on race, ethnicity, or other socially relevant groupings	We report split of race and ethnicity in our dataset. No separate analysis on race was performed as subpopulation bias was not in the scope of our work.
Population characteristics	The mean age of our dataset is 54.1 +- 16.6 years old. There were 6305 Females out of the 14214 studies included. All data was taken from adult patients.
Recruitment	N/A
Ethics oversight	The Cleveland Clinic Institutional Review Board waived the need to obtain informed consent for purposes of this study (IRB #22-1213).

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	Our training dataset was 11028 studies, our internal testing dataset was 2758 studies, and our external testing set was 428 studies. No specific sample size calculations were done as data used in the study was from routine clinical care.
Data exclusions	Only data from adult patients were included given most cardiac magnetic resonance imaging was done in adults. Only patients with both Cine and LGE imaging were included given these two exams provide the most value for CMR. The results are report in the body of the manuscript and show that replication of results are successful and reproducible. We have also used our model internally for classification of new CMR studies which we have found the accuracy to match our reported results.
Replication	We have two external validation datasets. 1) ACDC which is a public CMR dataset to enable readers to reproduce the work. 2) We validated our model on Cleveland Clinic Florida satellite sites. These sites are run independently of the Ohio sites, have different readers, different vendors, and different distribution of patient populations.
Randomization	We randomized training and test sets via random patient sampling.
Blinding	The investigators were blinded to group allocation. Many of the finalized disease labels were reused from other projects which the core investigators were not involved in. When training and test splits were created, we did not explicitly modify the training/testing sets to include or exclude any of them.

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

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☒	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Hazards: Improper application of AI could impact safe and effective treatment of patients undergoing CMR.

For examples of agents subject to oversight, see the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern.

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Precautions and benefits

Biosecurity precautions: N/A

Biosecurity oversight: N/A

Benefits: May provide low-cost, expedited interpretation of CMR images.

Communication benefits: Yes, there is minimal risk which may impact public safety because no direct intervention on patients is studied.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A