nature research

Corresponding author(s): NMED-A90984A

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	All data utilized in this study were electronically retrieved. SAS, Python, SQL and other core systems of our data warehouse were utilized.
Data analysis	SAS version 9.4 (Phreg and general descriptive analyses); Keras version 2.0.3 and TensorFlow 1.0.1 (CNN training); Python 2.7

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Computer code is available upon request. The raw patient data is not publicly available to institutional policy and human subjects approval.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study does not follow the traditional null hypothesis testing strategy that lends itself to power calculation (i.e., there is no null hypothesis for the algorithm). Instead, we abstracted all of the cases that met inclusion criteria and utilized measures of statistical precision. For sample sizes over 2000, which is this case for this study, precision of proportions (i.e., AUC, accuracy, sensitivity, etc) have precision <1 percentage point. Thus, the sample size of the study was sufficiently sized to support the research objectives.
Data exclusions	This study design required that the electrocardiogram (ECG) and echocardiogram (echo) be obtained within two weeks. As a result, 461,434 of the available patients with both an ECG and echo were excluded. Further eliminations were utilized to ensure that each patient was only present in the dataset once. This further eliminated 63,863 potentially eligible ECG-echo pairings. With the selection of the first observed paired assessment, a total of 97,829 were available for the primary analysis
Replication	The study design included separate training and test data. The test data was not used to determine model estimates. The replication of the findings from the training data was tested on this withheld data. In addition, a validation dataset was used to optimize model configuration as the model was being trained. This validation data was not utilized to assess the replication of the algorithm.
Randomization	This was a retrospective chart review study. There was no intervention or treatment assignment as a result. Therefore, randomization is not applicable.
Blinding	As noted in the randomization section above, there was no randomization or treatment assignment. Accordingly, blinding was not applicable to this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

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	This study population consisted of patients 18 years of age or older seen at Mayo Clinic in Rochester, MN. Eligible patients were required to have at least one digital, standard 10-second 12-lead ECG acquired in the supine position between January 1994 and February 2017 and at least one transhoracic echocardiogram obtained within 14 days of the ECG. The overall patient population had a mean age of 61.8+/- 16.5 years, and 7.8% of the population had a left ventricular ejection fraction of <35%
Recruitment	N/A - records based research