nature portfolio

Corresponding author(s): Pierre Elias, MD

Last updated by author(s): Mar 28, 2025

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	The dataset used to develop and evaluate the EchoNext model comprises over 1.2 million ECG-echocardiogram pairs from 230,318 patients across eight hospitals in the NewYork-Presbyterian health system. Due to institutional policies and patient privacy considerations, the full clinical dataset cannot be publicly released. However, a de-identified subset of 100,000 ECGs with paired echocardiographic labels has been made publicly available at {XXXXXXXXX}. Requests for access to additional data for academic or research purposes may be directed to the corresponding author and will be reviewed by the appropriate institutional committees. The Columbia Mini-Model was developed using the training and validation data splits provided in the publicly- released Columbia ECG dataset (Supplemental Table 9). The code is available at a GitHub repository at https://github.com/PierreElias/EchoNext_Validation. This repository is normally private but has been made public while paper is under review.
Data analysis	All statistical analyses were performed using Python version 3.8. EchoNext model version 4.0 was used. Reported performance is based on the public test set (Supplemental Table 11); performance was additionally assessed in the NYP Multicenter Cohort for direct comparison with the EchoNext model (Supplemental Table 10). As with EchoNext, this is a multi-task model that predicts both the composite SHD label and a component disease. Performance for each individual disease component is included in Supplemental Tables 10, 11.

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

To facilitate future research and innovation in this field, a de-identified 100,000-patient subset of the NYP Multicenter Cohort (specific to Columbia University Medical Center) will be released with publication of this study. This dataset will contain all input features needed to run the model (ECG waveforms, tabular ECG features, and patient demographics), as well as the corresponding disease labels derived from echocardiography. To provide a benchmark for future comparison, we split this dataset into train/validation/test and report the performance of the model within the public dataset. The dataset will be made available on PhysioNet as we have done with prior datasets, but due to Columbia University policies can't be made available before publication of the manuscript.

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 further evaluated model performance across NYP hospitals, clinical contexts, and with respect to patient age and race/ethnicity (Table 2). Across hospitals (AUROC range: 80%–87%; Central Illustration) and clinical settings (AUROC range: 79%–84%), the model exhibited stable, generalizable performance. Similarly, there were no clinically relevant differences in model performance by race, ethnicity, or sex. More details can be found in Table 1 and 2.
Reporting on race, ethnicity, or other socially relevant groupings	We further evaluated model performance across NYP hospitals, clinical contexts, and with respect to patient age and race/ethnicity (Table 2). Across hospitals (AUROC range: 80%–87%; Central Illustration) and clinical settings (AUROC range: 79%–84%), the model exhibited stable, generalizable performance. Similarly, there were no clinically relevant differences in model performance by race, ethnicity, or sex. More details can be found in Table 1 and 2.
Population characteristics	We curated a dataset comprising 1,245,273 echocardiogram-ECG pairs from 230,318 unique patients (≥18 years) collected between December 2008 and 2022 at one of eight NewYork-Presbyterian (NYP) affiliated hospitals (Figure 1). This dataset was designated as the NYP Multicenter Cohort. The data were split at a patient level into train (149,819 unique patients with 796,816 ECG-echocardiogram pairs), validation (35,780 unique patients with 35,780 ECG-echocardiogram pairs), and test (44,719 unique patients with 44,719 ECG-echocardiogram pairs) sets with patient characteristics described in Table 1. The performance of the model was tested in three external cohorts from Cedars-Sinai Medical Center (n=10,177 patients), the Montreal Heart Institute (n=10,862), and the University of California San Francisco Medical Center (n=6,106). SHD prevalence was higher in the external sites (54%, 52% and 46%, respectively) compared with SHD prevalence in the NYP cohort (36%). Despite large differences in disease prevalence and other patient demographic features compared with the model training population (Supplemental Table 7), EchoNext generalized well to these cohorts (AUROC 78-80% and AUPRC 78%-80%) with comparable ROC/PRC curves to individual hospitals within the NYP Multicenter Cohort (Central Illustration).
Recruitment	Not applicable.
Ethics oversight	This retrospective study was conducted with approval of the Columbia University and Weill Cornell Institutional Research Boards with waiver of patient consent.

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	No statistical method was used to pre-determine sample size. Instead, we included all eligible patients who met predefined inclusion criteria across eight hospitals between 2008 and 2023, resulting in 230,318 unique patients and over 1.2 million ECG-echocardiogram pairs. The large dataset size was selected to ensure robust model development and validation across diverse clinical settings and demographics. For the prospective trial, a 100-patient sample size was selected to assess real-world feasibility and model performance in a screening context. In the cardiologist reader study, 13 cardiologists each reviewed a block of 50 ECGs (totaling 3,200 interpretations), which provided sufficient data to compare diagnostic accuracy between human readers and the model.
Data exclusions	-Patients < 18 years of age at the time of ECG during the years 2008-2022 -Non-diagnostic quality ECGs

-Ventricular paced ECGs
-Patients who did not have an echocardiogram within 1 year of an ECG
-Echocardiograms with repaired or replaced heart valves
-Echocardiograms without a LVEF, a wall thickness measurement, and one relevant valve finding

Replication

In addition to the primary data partition, we performed a sensitivity analysis generating and comparing several distinct models based on variations in the training and validation sets. Specifically, we leveraged existing institutional internal data partitions to train three separate models: 'East' using data exclusively from Weill Cornell Medical Center, NYP-Brooklyn Methodist Hospital, NYP-Queens Hospital, and NYP-Lower Manhattan Hospital; 'West' using data exclusively from Columbia University Irving Medical Center, NYP-Westchester Hospital, NYP-CHONY, and NYP-Allen Hospital; and 'Blend' using a subset of the combined data from both East and West (Supplemental Table 3). A single test set was created with data from East and West to compare the various models. It is worth emphasizing that approximately half of this blended test set represents "external" data for the models trained on East or West data alone, offering an opportunity to assess generalizability compared with the model trained with Blended data. Details of the different data partitions used for training are provided in Supplemental Table 4; the 'Blend' set was sub-set from the available data to approximate training sample size.

Model performance results are shown in Supplemental Table 5 and Supplemental Figure 3. Overall, differences in performance were minimal across models, with differences in AUROC and AUPRC of 0.2% and 0.1%, respectively between West and Blended models. These findings support excellent external generalizability of models trained on these data for this task, as the performance drop with or without external test data was minimal.

Randomization

Data were split at a patient level into train (149,819 unique patients with 796,816 ECG-echocardiogram pairs), validation (35,780 unique patients with 35,780 ECG-echocardiogram pairs), and test (44,719 unique patients with 44,719 ECG-echocardiogram pairs) sets.

Blinding

Blinding was not applicable to this study because it involved retrospective analysis of routinely collected clinical data and automated model training using pre-labeled ECG-echocardiogram pairs.In the cardiologist reader study, ECGs were presented in randomized order, and cardiologists were blinded to echocardiographic outcomes. Therefore, while blinding was not necessary for model development, appropriate blinding measures were used during the human-reader comparison component.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration: N/A

Study protocol: *Note where the full trial protocol can be accessed OR if not available, explain why.*

Data collection: *Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*

Outcomes: Structural Heart Disease: The presence of LV ejection fraction (LVEF) ≤45%; maximum LV wall thickness ≥1.3 cm; moderate or severe right ventricular (RV) dysfunction; pulmonary hypertension (PASP ≥45mmHg or TR jet velocity ≥3.2 m/s); moderate or severe aortic stenosis (AS), aortic regurgitation (AR), mitral regurgitation (MR), tricuspid regurgitation (TR), or pulmonary regurgitation (PR); or a moderate or large pericardial effusion

Plants

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