Copyright © 2025, Massachusetts Medical Society. All rights reserved.

This is an Author Accepted Manuscript, which is the version after external peer review and before publication in the Journal. The publisher's version of record, which includes all NEJM AI editing and enhancements, will be available at https://ai.nejm.org/doi/full/10.1056/AIoa2500164.

A foundation transformer model with self-supervised learning for ECG-based assessment of cardiac and coronary function

Jonathan B. Moody[1,6], Alexis Poitrasson-Rivière[1], Jennifer M. Renaud[1], Tomoe Hagio[1], Fares Alahdab[2], Mouaz H. Al-Mallah[3], Michael D. Vanderver[1], Sascha N. Goonewardena[4,5], Edward P. Ficaro[1,4], and Venkatesh L. Murthy[4,6]

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

[1] INVIA, LLC, Ann Arbor, MI, USA

[2] Department of Biomedical Informatics, Biostatistics, Epidemiology, Department of Cardiology, University of Missouri, Missouri, USA

[3] Houston Methodist DeBakey Heart and Vascular Center, Houston, TX USA

[4] University of Michigan, Division of Cardiovascular Medicine, Department of Internal Medicine, Ann Arbor, MI, USA

[5] VA Ann Arbor Health System, Ann Arbor, MI, USA

[6] Corresponding authors: jmoody@inviasolutions.com, vlmurthy@med.umich.edu

Abstract

Background: The wide availability of labeled electrocardiogram (ECG) data has driven major advances in artificial intelligence (AI)-based detection of structural and functional cardiac abnormalities and thus ECG-based diagnosis. However, many critical, high value clinical diagnostic applications, such as assessing myocardial ischemia and coronary microvascular dysfunction, remain underserved due to the limited availability of labeled datasets. We developed a self-supervised ECG foundation model and demonstrate how this approach can overcome this limitation.

Methods: A modified vision transformer model was pretrained using a large database of unlabeled ECG waveforms (MIMIC-IV-ECG, N=800,035). The model was then fine-tuned using smaller databases that included high-quality labels derived from positron emission tomography (N=3,126) and clinical reports (N=13,704) for 12 clinical, demographic, and traditional ECG prediction tasks. Diagnostic accuracy and model generalizability were evaluated across five additional cohorts including the publicly available PTB-XL and UK Biobank databases and labels from cardiac magnetic resonance imaging (MRI) and single photon emission computed tomography (SPECT).

Results: Diagnostic performance varied across tasks with area under the receiver operating characteristic curve (AUROC) ranging from 0.763 for detection of impaired myocardial flow reserve (MFR < 2) to 0.955 for impaired left ventricular ejection fraction (LVEF < 35%). Self-supervised learning (SSL) pretraining greatly improved diagnostic accuracy in 11 of the 12 prediction tasks compared to conventional *de novo* supervised training. The model retained strong performance across three external and two internal cross-modality databases, with AUROC ranging from 0.771 for impaired MFR to 0.949 for impaired LVEF.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Conclusion: This versatile ECG foundation model demonstrates that SSL pretraining enhances diagnostic accuracy and generalizability across diverse cardiac diagnostic applications. By enabling effective learning from limited labeled data, this approach supports AI development for complex but clinically critical tasks, such as detecting myocardial ischemia and coronary microvascular dysfunction, where high-quality labels are costly and scarce.

Description

This study introduces a self-supervised ECG transformer foundation model that achieves high diagnostic accuracy across multiple cardiac prediction tasks. By enabling robust performance with limited labeled data, the approach expands AI's potential in complex diagnostic areas such as myocardial ischemia and coronary microvascular dysfunction.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Introduction

Nearly 1 in 17 emergency department visits in the United States are for chest pain, representing almost 8 million encounters annually.[1] Chest pain also drives nearly 6 million outpatient clinic visits per year.[2] Cardiac stress testing is the dominant diagnostic approach for evaluating such patients.[3] Several noninvasive testing modalities provide important diagnostic measures of cardiac and coronary function, albeit with varying accessibility and cost (Figure 1). AI models of electrocardiograms (ECGs) have excelled at automating ECG interpretation and show promise as low-cost diagnostic tools.[4] Because of the wide availability of ground truth annotations from clinical ECG and echocardiographic findings, supervised learning using large, labeled databases has been the dominant paradigm for nearly all of these applications.[5]

ECGs can also detect myocardial ischemia and infarction, providing insight into perfusion and potassium balance, which inform the management of cardiovascular disease. Positron emission tomography (PET) myocardial perfusion imaging (MPI) is the gold standard for accurate characterization of ischemia and coronary microvascular dysfunction (CMVD).[6,7]

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Such testing enables quantitative cardiac and cardiovascular measures — most importantly, myocardial flow reserve (MFR), which has independent diagnostic and prognostic utility for known or suspected coronary artery disease.[8–10] However, PET MPI is costly and not widely available,[11] limiting the availability of data derived from quantitative PET measurements. Consequently, AI-enhanced approaches in this arena have been very limited.

Foundation models and self-supervised learning (SSL) have enabled paradigm shifts across many AI domains.[12] Foundation models are trained on broad data that can be adapted to a range of downstream tasks.[13] SSL enables such models to learn powerful task-agnostic representations from large unlabeled databases, allowing for efficient fine-tuning using smaller task-specific annotated databases.[14] This approach underlies many state-of-the-art large language and computer vision models.[15,16]

Despite these strengths, foundation models and SSL methods have yet to be widely applied to ECG signals. A related AI innovation, the transformer architecture,[17] now standard in

natural language processing,[18] excels at modeling long-range dependencies in sequential data but has rarely been applied to ECGs.[19,20]

In this work, we developed an SSL foundation model based on a modified vision transformer,[21] and pretrained it using a generative pretext task in a large database of 800,035 unlabeled ECG waveforms. The model was then fine-tuned in smaller labeled databases for 12 demographic and clinical prediction tasks across 3 major dimensions of cardiology: myocardial function, coronary perfusion, and cardiac rhythm. Four tasks focused on assessing myocardial ischemia and coronary microvascular and vasomotor dysfunction (CMVD), which have been particularly inaccessible with ECG-AI models. We demonstrate the efficacy of the SSL approach by comparing the foundation model with an analogous model developed by conventional *de novo* supervised training. The models were evaluated across three external and two internal cross-modality databases for diagnostic accuracy and generalizability. Results show that combining SSL with a transformer architecture enables efficient training and robust diagnostic performance of high-value, clinically impactful ECG-based AI tools with limited labeled data.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Methods

Data sources

Seven ECG databases were used for model development, validation, fine-tuning, and testing Table 1. The first was the publicly available MIMIC-IV-ECG repository v1.0[22,23] of standard 12-lead resting ECGs of 10 seconds duration. These waveforms were used as unlabeled data for pretraining the SSL foundation model.

Three ECG databases were obtained from MPI registries at the University of Michigan Frankel Cardiovascular Center. These consisted of all patients who underwent clinically indicated stress testing with ECG and concurrent noninvasive MPI with either pharmacologic stress positron emission tomography/computed tomography (PET/CT) (UM-PET) or single photon emission computed tomography (SPECT) using pharmacologic (UM-phSPECT) or exercise (UM-exSPECT) stress. The UM-PET database was used for model development, internal validation, and fine-tuning and the two SPECT databases for model evaluation and internal, cross-modality testing.

Quantitative measurements of gold standard PET-MPI data were used to generate ECG labels for task-specific fine-tuning of the foundation model. The PET data were randomly split at the patient level (not study level) into training, validation, and holdout test subsets with the ratios ~60:20:20% (N=3126/1041/1031). All three Michigan cohorts were reviewed to remove any patient-level overlap. Patients who had undergone both PET/CT and SPECT/CT exams were included only in the UM-PET cohort and excluded from SPECT cohorts to avoid label leakage. Only the earliest evaluable exam was included for patients with more than one MPI exam. Patients were excluded if they had a history of heart transplantation, missing or uninterpretable image data, or missing demographic or hemodynamic data. All patient data was de-identified and informed consent was waived under an exemption from the University of Michigan Institutional Review Board.

The fifth database was a cohort of clinical patients undergoing routine PET-MPI stress testing at the Houston-Methodist hospital and was used for external testing of PET-MPI prediction tasks (HM-PET).

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

The sixth database was the publicly available PTB-XL ECG database.[24,25] The predefined folds of this data were used to fine-tune and evaluate the SSL foundation model for automatic ECG interpretation and diagnosis tasks. The entire dataset was also used for externally testing the SSL model fine-tuned on the UM-PET database for the three demographic prediction tasks.

The last database was the publicly available UK Biobank database[26] which was used for external testing of left ventricular ejection fraction (LVEF) and demographic tasks.

Noninvasive stress testing

Routine rest and stress cardiac ^{82}Rb PET exams were performed in accordance with clinical guidelines for MPI testing[27] and measurement of myocardial blood flow (MBF)[28,29] as previously described.[30]

Cardiac ^{99m}Tc-sestamibi SPECT exams were performed in accordance with guidelines[31] as previously described.[32] LVEF[33] and stress total perfusion deficit (TPD)[34] were routinely estimated for all imaging exams. Pharmacologic stress was induced in the UM-PET

database with intravenous bolus administration of regadenoson (0.4 mg) and in the UM-phSPECT database with either regadenoson alone or in combination with low-level treadmill exercise. In the UM-exSPECT database, stress was induced by treadmill exercise using a Bruce, modified Bruce, or Cornell protocol. Heart rate, systolic, and diastolic blood pressure were monitored continuously during imaging.

Twelve-lead ECG waveforms of 10 second duration and 500 Hz sampling rate were recorded at baseline immediately before stress testing and at 1-minute intervals during stress (in supine position for pharmacologic or combination pharmacologic/low-level exercise stress) or once during each exercise stage (for exercise stress). Both rest and stress ECG waveforms accompanying each PET MPI exam were used for ECG-AI model development.

Data annotations

ECG annotations were derived for stress and resting ECGs acquired simultaneously with PET-MPI (nine prediction tasks) or SPECT-MPI (five tasks); for resting ECGs acquired on the same day as cardiac magnetic resonance imaging (MRI) in the UK Biobank data (four tasks); and for resting ECGs in the PTB-XL data (six tasks) Table 2.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Absolute MBF in ml/min/g of tissue was measured from 30-frame dynamic PET images at rest and stress with a clinically validated, commercially available one-tissue compartmental model, and MFR was computed as the ratio of stress over rest MBF and averaged globally over the entire left ventricular myocardium.[35,36] Stress TPD was computed from relative myocardial perfusion images as previously described.[37]

Left ventricular volumes were measured from 16-frame ECG-gated myocardial perfusion images using a clinically validated, commercially available algorithm, and LVEF was computed as one minus the left ventricular (LV) volume ratio at systole over diastole.[33] LVEF reserve was computed as the difference between stress and rest LVEF.[38] In the UK Biobank database, LVEF measurements derived from cardiac MRI were obtained from the publicly available database.[39,40]

These six tasks were treated as individual single label binary classifications with appropriate thresholds for continuous measures obtained from the literature and clinical experience (MFR < 2, stress MBF < 1.5 ml/min/g, rest MBF > 1.2 ml/min/g,[28] stress TPD > 5%,[41] LVEF < 35%,[42] LVEF reserve < -5%[38]).

Three auto ECG interpretation tasks were treated as a multilabel classification tasks with three categories of ECG statements as defined in the PTB-XL dataset: 44 diagnostic statements, 12 rhythm statements, and 19 form statements.[24]

For the three demographic prediction tasks, age and BMI were treated as continuous regression tasks, and sex as a binary classification task with labels determined by patient records.

Deep learning model

Model development

The primary inputs for the SSL foundation model were raw 12-lead ECG waveforms of 10 second duration. A modified vision transformer neural network[21] was developed by adapting the initial patch embedding layer for one-dimensional, multichannel input (1dViT) (Supplementary Figure S1 (a)). Waveform patches of fixed 100 millisecond duration corresponding to a patch size of 50 were flattened, linearly projected into an embedding space of dimension 800, combined with a trainable class token, and summed with trainable positional encoding.[21] A series of 12 transformer encoder blocks followed, each incorporating pre-layer normalization,[43] 8 self-attention heads, and a multi-layer perceptron with GELU activation[44] and hidden dimension of 3200 (Supplementary Figure S1 (a)). The final classifier head consisted of global average pooling, layer normalization, and a linear classifier.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

The number of input ECG leads was reduced from 12 to 8 (2 bipolar and 6 precordial) since the 3 augmented unipolar leads are linear combinations of the bipolar leads, and by Einthoven's triangle equation only 2 of the 3 bipolar leads are linearly independent.[45] A fusion model for dual input using resting and stress ECG waveforms was constructed by concatenating the outputs of two pretrained 1dViT backbones with classifier head removed

(Supplementary Figure S1 (b)) and combining with a final linear classifier. This fusion model was fine-tuned or trained from scratch (see Model fine-tuning below) using end-to-end learning using both stress and rest ECG input for each patient. The 1dViT backbone contained 92.7 million trainable parameters and, except for the modified 1D patch embedding and patch size (50), its architecture is similar to a 2D ViT-B/16.[21] All models were developed using the PyTorch framework (v2.1.1).[46] Our code is available at https://github.com/4dm-labs/ecgflow and pretrained model checkpoints are available at https://huggingface.co/4dm-labs.

ECG preprocessing

Waveforms from the University of Michigan and Houston-Methodist Hospital underwent routine filtering for powerline interference and low-frequency baseline wander[47] using vendor software on the stress ECG carts. Waveforms from PTB-XL and UK Biobank databases, which had no prior vendor preprocessing, underwent comparable in-house preprocessing using zero phase distortion bandpass (0.01 - 150 Hz)[48] and notch filtering (0.001 Hz). A small number of waveforms from the MIMIC-IV-ECG databases had missing sample points which were imputed using the mean. Each waveform lead was additionally normalized to zero mean and unit standard deviation (SD) using the grand mean and SD of either the MIMIC-IV-ECG database during SSL pretraining or the UM-PET ECG training database during fine-tuning. The grand mean (SD) in the MIMIC-IV-ECG training split (N=780,035) for 8 leads was: 0.0099 (0.2166) mV. This was similar to the grand mean (SD) in the UM-PET training split (also in physical mV units): 0.0153 (0.2333).

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

SSL model pretraining

We adapted for 1D ECG data a particularly simple generative SSL methodology called the masked auto-encoder.[14,49] For SSL pretraining, the 1dViT backbone without the classifier head was paired with a small transformer decoder to perform a pretext task of masked signal modeling.[50,51] The decoder had eight transformer blocks similar to the 1dViT encoder but with embedding dimension of 256 and 16 self-attention heads.[51] Embedded input patches were randomly masked with a masking ratio of 60%, and only unmasked patches were input to the encoder.[51] Before entering to the decoder, encoded patches were

combined with a trainable token for each masked patch and summed with positional encoding. The input waveform was reconstructed at the level of raw time steps with an L_2 loss.[51] Our pretraining method closely follows generative SSL methods from the literature.[21,50,51]

The model was trained using the AdamW optimizer[52] with an initial learning rate of 1e-5, linear warmup for five epochs to a peak of 1e-3 followed by a cosine decay schedule.[53] Pretraining was performed for 300 epochs with a batch size of 384.

Model fine-tuning

The SSL pretrained 1dViT encoder was subsequently fine-tuned on labeled data for each prediction task. In addition to the fusion 1dViT for dual stress-rest ECG input, a single 1dViT model was also fine-tuned for rest-only ECG input to evaluate the additional performance provided by using stress ECGs. Only the classifier head and top four transformer blocks were trained during fine-tuning, while freezing the remaining eight encoder blocks. A binary cross-entropy loss function was utilized for multi- and single-label tasks and mean squared error loss for continuous regression tasks. We used the AdamW optimizer with linear warmup of the learning rate from 1e-5 to a peak of 5e-5 followed by cosine decay. Model and training hyperparameters were set using suggested values from the literature.[21,54] Label smoothing regularization[55] was tuned to optimize model calibration[56] by minimizing the mean absolute difference and 90th percentile absolute difference between observed and predicted probabilities — integrated calibration index (ICI) and E90.[57] For the auto ECG interpretation tasks, positive label weights were used to compensate for label imbalance during training and fine-tuning. Fine-tuning was performed to maximize area under the receiver operating characteristic curve (AUROC) using 100 epochs with a batch size of 256. Training iterations were stopped early when validation AUROC failed to improve for ten consecutive epochs. For comparison, the same model without SSL-pretraining was trained and validated *de novo* from random initialization with the same labeled data and training setup, except the peak learning rate during linear warmup was set to 1e-3 as used for SSL pretraining.[54] In the same way, fine-tuning of the SSL pretrained model and supervised *de novo* training for the auto ECG interpretation tasks was performed using the PTB-XL predefined training folds.[24]

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Model evaluation

Feature maps generated at the last transformer block were isolated and projected into two dimensions for cluster analysis using the LocalMAP refinement of the PaCMAP dimension reduction algorithm[58,59] (Supplementary Figures S14 – S17).

Model discrimination of binary classification tasks was assessed using ROC curves, diagnostic accuracy, AUROC, and macro averages for multi-label tasks. Model calibration for these tasks was assessed graphically with calibration curves of predicted and observed probabilities, and quantitatively using ICI and E90. Continuous regression tasks were evaluated with scatter and Bland-Altman plots of true and predicted values, mean absolute error (MAE), and R^2. All model evaluations in the UK Biobank and PTB-XL databases used the rest ECG single input 1dViT model since appropriate stress ECGs were not available in these databases (limited 4-lead exercise ECGs that are available in UK Biobank were not adapted here).

Label efficiency was evaluated for five tasks over two databases: LVEF and MFR (UM-PET), and the three auto ECG interpretation tasks (PTB-XL). The pretrained SSL model was fine-tuned over a series of randomly sampled data subsets of decreasing size. For comparison, the same model was also trained *de novo* from random initialization using the same hyperparameter settings described above. Each model was then evaluated with holdout data and the entire process was repeated five times with different random seeds.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Statistical methods

For diagnostic evaluation, the DeLong test[60] was used to compare single-label AUROCs, McNemar test[61] for comparison of sensitivity and specificity, and Leisenring test[62] for comparison of positive and negative predictive values. Bootstrap methods were used to compare multilabel models.[63] To facilitate comparison between ROC curves, a threshold was selected at the point on the curve that minimized the distance to the point of perfect discrimination (sensitivity=1, specificity=1). Subgroup analysis of model performance was performed, stratifying by age (< 60 y), BMI (< 30 kg/m^2), sex, diabetes, dyslipidemia, hypertension, known coronary artery disease (CAD), LVEF (< 50%), and stress TPD (< 5%).

Continuous variables are summarized as mean±SD or median [1st – 3rd quartiles]. Welch's unequal variances t-test or Wilcoxon rank sum tests were used as appropriate for comparisons of continuous parameters, and χ-squared tests were used to compare categorical variables. Two-sided p-values less than 0.05 were considered statistically significant.

Results

Patient populations

The MIMIC-IV-ECG repository provided 800,035 diagnostic resting ECGs acquired in 161,352 patients. For SSL pretraining, ECG data were split into training/validation subsets 780,035/20,000 (97.5%/2.5%) and no other patient data were used. In the UM-PET database, after excluding 259 heart transplant patients and 4 PET/CT patients due to uninterpretable PET data or missing demographic data, 5,198 patients were available for ECG-AI model development and testing: 3,126 (training or fine-tuning), 1,041 (validation), and 1,031 (holdout test) (Supplementary Figure S2). The sizes of evaluation databases are listed in Table 1.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Baseline characteristics

Baseline characteristics are shown in Table 3 stratified by patient cohort. As expected, the PET MPI cohorts for model derivation and holdout test had nearly identical demographics and risk factor profiles. However, the pharmacologic stress SPECT MPI testing cohort differed meaningfully from PET MPI cohorts. SPECT patients were older, with lower rates of obesity, diabetes, and history of MI, and higher rates of hypertension, hyperlipidemia, and mortality. Baseline characteristics of the exercise stress SPECT MPI cohort are compared in Supplementary Table S1. The number of positive cases for each task and cohort (overall and within subgroups) is shown in Supplementary Figures S6 - S10. In particular, the prevalence of impaired LVEF was very low in the UM-exSPECT (0.5%) and UK Biobank (0.3%) cohorts.

SSL pretraining

Figure 2 compares ECG-AI model performance with and without SSL pretraining for nine tasks evaluated in the external HM-PET database. SSL pretraining produced significant AUROC increases from +0.055 to +0.103 for all binary classification tasks except prediction of impaired LVEF reserve (+0.006, $p = 0.850$; see Supplementary Tables S2, S3, and S4 for complete summaries including point estimates and 95% confidence intervals for each evaluation metric). In addition, R^2 was increased and MAE reduced for continuous regression tasks (Supplementary Table S2). Additional testing of these same tasks evaluated in the UM-PET holdout database yielded similar diagnostic performance and SSL benefits (Supplementary Figure S3, Supplementary Table S4). Strong diagnostic performance was maintained in internal, cross-modality tests evaluated in the pharmacologic (UM-phSPECT) and exercise (UM-exSPECT) stress SPECT cohorts, and external tests in the UK Biobank and PTB-XL databases (Figure 3) (labels were not available for every task).

Supplementary Table S3 and Supplementary Figure S4 compare ECG-AI model performance for three auto ECG interpretation tasks in the PTB-XL database. The increased performance with SSL pretraining was maintained in terms of macro AUROC across all tasks (+0.035 to +0.129, $p < 0.0097$), as well as macro Youden index for two of the three tasks (+0.100 to +0.219, $p < 0.0089$).

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Label efficiency of the SSL-pretrained model is compared to *de novo* training in Figure 4 for four tasks across two databases. Label efficiency varied by task and database but generally improved with SSL pretraining. The *de novo* model AUROC tended to plateau at < 100% of the training data while the SSL-pretrained model continued to increase. In all four tasks the SSL model reached the highest AUROC achieved by the *de novo* model with far less than 100% of the training data.

Two examples of model calibration are shown in Supplementary Figure S5. Model calibration was improved in terms of ICI and E90 after tuning label smoothing regularization.

Subgroup analysis demonstrated broadly consistent model performance over multiple tasks and evaluation databases (Supplementary Figure S6 to Supplementary Figure S10).

Benefits of stress ECG

Some performance increases were observed for the dual input (stress-rest ECG) 1dViT model compared to the single-input (rest-only ECT) model. However, they tended to be more modest in magnitude and scope than that of SSL pretraining. AUROC improved for the stress MBF task (rest-only ECG 0.841 95% CI [0.812 - 0.870] vs. stress-rest ECG 0.866 95% CI [0.840 - 0.892], $p = 0.0052$) and stress TPD (rest-only ECG 0.740 [0.707 – 0.772] vs. stress-rest ECG 0.762 [0.730 – 0.794], $p = 0.0411$). However, changes in sensitivity and specificity using stress-rest ECGs were mixed (sensitivity: MFR +0.061 [0.0256 – 0.0956]; stress MBF +0.005 [-0.0371 – 0.0464]; LVEF +0.062 [0.0056 – 0.1183]; specificity: MFR -0.047 [-0.0754 – -0.0179]; stress MBF +0.056 [0.0334 – 0.0793]; LVEF -0.053 [-0.0697 – -0.0371]). Note that all external evaluations with UK Biobank and PTB-XL databases (including SSL vs. non-SSL) used the single-input rest-only model as stress ECGs were not available.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Discussion

We applied a label efficient SSL approach to implement a foundation model capable of performing multiple clinically relevant ECG-based prediction tasks with limited labeled data. The model was pretrained with a self-supervised learning method on a large publicly available database of unlabeled ECG data and was then fine-tuned on smaller datasets with labels derived from PET, patient demographics, and clinical ECG reports targeting 12 prediction tasks. The model was validated in five additional datasets that included labels derived from two additional imaging modalities (MRI and SPECT). Compared with *de novo* supervised training, the SSL paradigm consistently improved the diagnostic accuracy of 11 of the 12 prediction tasks investigated. Importantly, this approach may facilitate the development of deep learning-based, low-cost testing strategies for more complex cardiac diagnostics previously thought to be inaccessible due to the need for extremely large training datasets.

ECG-AI prediction tasks

Out of our 12 ECG prediction tasks, 7 have been previously studied. ECG-AI prediction of LVEF was reported by Attia et al.[42] and has been externally validated in multiple prospective cohorts.[64] In our small, PET-derived LVEF dataset, we observed remarkably high diagnostic accuracy after *de novo* supervised training, which was significantly improved with SSL pretraining and fine-tuning, and remained consistently high in external and internal cross-modality testing.

ECG-AI prediction of age, BMI, and sex has been reported by several groups,[63,65–68] under the hypothesis that differences between predicted and actual values reflect general cardiovascular or cardiometabolic health, or hormonal status.[69] Again, we observed quantitative and diagnostic accuracy comparable to prior studies,[63,67,68] including in two publicly available datasets (PTB-XL and UK Biobank), and the accuracy was further improved by SSL pretraining. Similarly, the SSL foundation model fine-tuned in the PTB-XL database also increased diagnostic accuracy for three auto ECG interpretation tasks compared with *de novo* training.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Five of our prediction tasks leveraged rest and stress PET-derived labels that have not been previously studied with deep neural network ECG modeling. Stress TPD is a relative measure of myocardial perfusion that indicates the extent and severity of focal obstructive epicardial disease.[34] Stress and rest MBF and MFR (MFR = stress MBF divided by rest MBF) are measures that provide complimentary information and add independent, incremental diagnostic and prognostic value to stress TPD measurements.[70,71] LVEF reserve provides additional information on the presence of high-risk multi-vessel disease.[38,72] Together these PET-derived measures provide comprehensive, state-of-the-art assessment of left ventricular myocardial and coronary function in patients with known or suspected coronary artery disease. Additionally, since stress and rest ECGs were acquired simultaneously with each PET scan, these measures provide high-quality data annotations for ECG model training and fine-tuning.

In addition to external testing of the ECG-AI model with cardiac PET data from the Houston-Methodist Hospital, further testing was performed in four other databases (UK

Biobank, PTB-XL, UM-phSPECT, and UM-exSPECT). Although the UM-phSPECT and UM-exSPECT databases were collected at the same institution as the UM-PET model derivation database, these cohorts actually represent out-of-distribution datasets with respect to the UM-PET cohort. There were significant differences between these cohorts (Table 3 and Supplementary Table S1) in clinical referral to stress testing; imaging and stress modality; patient demographics, symptoms, and cardiovascular risk factors; resting and stress hemodynamics during image and ECG acquisition; LVEF and incidence of impaired LVEF (< 35%); stress TPD and incidence of abnormal stress TPD (< 5%); and clinical outcomes (Supplementary Figure S12). The 1dViT feature maps also exhibited distinct global structures among these cohorts (Supplementary Figures S16 – S17). Labels were not available for all tasks due to the retrospective nature of the data, and labels for the LVEF and stress TPD tasks were derived from modalities (MRI and SPECT) with significantly different image resolution, noise properties, and contrast compared to the model derivation database (PET). Nevertheless, the SSL pretrained model retained strong diagnostic performance similar to that of UM-PET holdout testing.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

SSL pretraining

A number of different SSL pretraining approaches have been described in the computer vision literature which may be broadly classified as contrastive or generative.[73] Although these methods have often performed similarly in the image recognition literature, generative approaches have been reported to perform better in the context of multichannel time-series modeling[49] and 2D retinal image modeling.[14] Our generative approach is analogous to that of Zhou et al.[14] which similarly showed strong performance over several prediction tasks. The regularity of these types of data across patients may make contrastive methods less suitable for pretraining, although further investigation is warranted.

Impact of stress ECG

Four of our tasks address prediction of ischemia-related parameters (MFR, stress MBF, stress TPD, and LVEF reserve) that involve pharmacologic or exercise stress testing and provided both resting and stress ECGs. Although we expect stress perturbations to add important additional information to model predictions for these tasks, we only observed a

modest increase in task performance compared to rest-only ECG input. In this retrospective observational study, we were unable to definitively identify a mechanism for the relatively high informational content of resting ECGs compared to stress. Stress ECGs are often subject to more noise and artifacts than resting ECGs and routinely undergo more aggressive filtering which may suppress some useful signals. Further, our SSL pretraining only included resting ECGs from the MIMIC-IV-ECG database. We aim to explore this question in future studies by including stress ECGs in SSL pretraining and further analysis of feature map clustering.

Related work

Very few ECG-AI models have been reported that address angina patients undergoing stress testing. To our knowledge, our ECG foundation model is the first deep neural network that uses both rest and stress ECGs to detect multiple aspects of impaired myocardial perfusion, confirming that tissue- and molecular-level pathophysiology results in characteristic electrophysiologic changes detectable in surface ECG tracings. Ahmad and colleagues[74] performed a ground-breaking study on a similar question in patients with suspected CMVD. Using a range of machine learning approaches with resting ECG data, they built prediction models to detect impaired coronary flow reserve, a measure analogous to MFR and identified by invasive coronary vasomotor reactivity testing. Although such invasive testing is precise, it is not routinely performed and can be subject to substantial referral bias. Because of the limited sample size and lack of stress perturbation data, the accuracy of their prediction models was only moderate.[74] Alahdab et al.[75] recently reported an ECG machine learning model based on a PET-derived labels to predict MBF and MFR from resting ECGs. In an independent cohort of patients who underwent SPECT MPI, they demonstrated that ML-predicted MFR was significantly associated with risk of major adverse cardiac events. Our results extend these earlier findings and confirm the utility of ECGs for AI-enhanced assessment of myocardial ischemia and CMVD.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Several previous studies have reported high-performing foundation models based on contrastive, token-based SSL-pretraining or conventional supervised learning.[76–78] These have primarily targeted traditional ECG interpretation, LVEF, and cardiac biomarker tasks and have benefited from very large, private, labeled ECG databases. By contrast, our choice

of tasks was largely driven by the desire to leverage high-value cardiac and coronary diagnostics using much smaller, less accessible labeled databases. Vaid et al.[20] recently reported a foundation model based on the BEiT transformer framework which also uses a token-based self-supervised pretraining method with unlabeled data. However, the BEiT framework operates on 2D images of ECGs rather than the raw waveform signals as used in our model. In 2D benchmark comparisons reported by He et al.[51] our SSL pretraining method performed somewhat better on downstream tasks than that used by BEiT.

Clinical implications

Many successful ECG-based AI prediction models have been reported for such tasks as automatic ECG interpretation,[4,5,63,79] detection of left ventricular (LV) and right ventricular (RV) systolic dysfunction,[42,64,80] and demographic inference.[68,69] These studies have been enabled by the accessibility of very large retrospective datasets derived from general ECG or echocardiography populations, and the models are robust in diverse external populations as observed in our study.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

In contrast, the smaller stress testing populations considered in this study are underrepresented in the ECG-AI literature, reflecting the difficulties of training strong AI models that address important diagnostic challenges for these patients. These smaller populations are characterized by a high incidence of obesity, diabetes, and other cardiovascular risk factors and comorbidities which are strongly associated with CMVD development.[7] CMVD causes myocardial ischemia and symptomatic angina without obstructive coronary disease[81,82] and is associated with markedly increased rates of adverse cardiac outcomes.[10,83] Despite high prevalence and prognostic relevance,[10] CMVD is difficult to diagnose with traditional clinical tools, and many patients remain undiagnosed or have presumptive diagnoses without confirmation. Advanced diagnostic measures that can identify CMVD include quantitative coronary or myocardial flow reserve as measured by invasive vasomotor reactivity testing or noninvasive stress imaging with quantitative PET MPI,[9] transthoracic doppler echocardiography, or cardiovascular MRI.[84]

These advanced techniques are not widely available and are substantially more costly than standard cardiovascular testing methods, such as invasive coronary angiography and

exercise stress ECG, which cannot quantify MFR. Among the advanced tests, quantitative PET MPI is considered the noninvasive gold standard for assessing CMVD,[85] and likely provides the highest quality data labels for AI model training. Although guidelines broadly recommend PET testing for a range of serious clinical conditions,[88] its accessibility remains limited.[11] In resource-constrained environments, ECG-AI predicted MFR could provide a cost-effective initial test to direct patients who would benefit most from quantitative PET MPI testing. Further, ECG-AI MFR could be useful as an adjunct to other more accessible stress testing modalities such as stress SPECT MPI, MRI, stress echocardiography, and exercise ECG, providing valuable additional information on CMVD phenotyping.

Limitations

The present study did not evaluate the diagnostic performance of ECG-AI models for detection of obstructive CAD. Although inaccurate assessment of "balanced" multivessel CAD is a long-standing limitation of standard SPECT MPI,[89] PET MPI enables much more accurate detection.[90,91] We speculate that ECG-AI MFR impairment could potentially improve the diagnostic accuracy of MPI in the presence of high-risk left main or multi-vessel obstructive CAD, as has been shown for quantitative PET MFR in prior studies.[92–94]

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Global left ventricular PET MFR does not distinguish CMVD from myocardial perfusion impairments due to focal epicardial or diffuse CAD. Our recent work has shown that regional quantitative PET measures combined with relative perfusion and global MFR can be highly effective at assessing the additive risk of diffuse or microvascular disease.[94,95] Further work will be necessary to integrate these results with the present ECG-AI model predictions to further characterize CMVD.

Although we provide benchmark comparisons with a variety of models from the literature (Supplementary Tables S5-S9), in this study SSL benchmarking with other architectures was not performed.

Conclusion

Our ECG foundation model using SSL and fine-tuning for multiple clinical tasks with limited labeled data demonstrated strong diagnostic performance in five independent databases.

The model is uniquely capable of identifying impaired MFR, a hallmark of CMVD, demonstrating that a pretrained foundation model can enable AI applications that are currently underdeveloped due to costly and scarce high-quality labeled data.

Acknowledgements

The authors acknowledge the Regents of the University of Michigan for the use of de-identified clinical data for this study. This research has been conducted using the UK Biobank Resource under application number 234540.

Funding

SNG is supported by VA MERIT grant 1I01CX002560, NIH/NHLBI R01HL150392, and the Taubman Medical Research Institute (Wolfe Scholarship). VLM is supported by grants R01AG059729 from the National Institute on Aging, U01DK123013 from the National Institute of Diabetes and Digestive and Kidney Disease, and R01HL136685 from the National Heart, Lung, and Blood Institute as well as the Melvyn Rubenfire Professorship in Preventive Cardiology.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Author Declarations

JBM, APR, JMR, TH, and MDV are employees of INVIA. JMR is a consultant for Jubilant Radiopharma and receives royalties from the licensing of FlowQuant software. EPF is a stockholder in INVIA. VLM has received research grants and consulting fees from Siemens Healthineers and serves as a scientific advisor for Ionetix and owns stock options in the same. He owns stock in GE and Cardinal Health and has received payments for consulting from INVIA. FA, MHA, and SNG have nothing to declare.

Ethics Approval

This is an observational study, and informed consent was waived under an exemption from the University of Michigan Institutional Review Board.

References

1. Cairns C, Kang K. National Hospital Ambulatory Medical Care Survey: 2021 Emergency Department Summary Tables [Internet]. 2021. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHAMCS/doc21-ed-508.pdf

2. Santo L, Kang K. National Hospital Ambulatory Medical Care Survey: 2019 National Summary Tables [Internet]. National Center for Health Statistics (U.S.); 2019 [cited 2024 Mar 31]. Available from: https://stacks.cdc.gov/view/cdc/123251

3. Gulati M, Levy PD, Mukherjee D, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain. Journal of the American College of Cardiology [Internet] 2021 [cited 2021 Oct 29];Available from: https://www.jacc.org/doi/10.1016/j.jacc.2021.07.053

4. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol [Internet] 2021 [cited 2023 Jun 26];18(7):465–78. Available from: https://www.nature.com/articles/s41569-020-00503-2

5. Li J, Aguirre AD, Junior VM, et al. An electrocardiogram foundation model built on over 10 million recordings. NEJM AI 2025;2(7):AIoa2401033. Available from: https://ai.nejm.org/doi/10.1056/AIoa2401033

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

6. Schindler TH, Schelbert HR, Quercioli A, Dilsizian V. Cardiac PET imaging for the detection and monitoring of coronary artery disease and microvascular health. J Am Coll Cardiol Img [Internet] 2010 [cited 2011 May 13];3(6):623–40. Available from: http://imaging.onlinejacc.org/cgi/content/abstract/3/6/623

7. Del Buono MG, Montone RA, Camilli M, et al. Coronary microvascular dysfunction across the spectrum of cardiovascular diseases. Journal of the American College of Cardiology [Internet] 2021 [cited 2023 Aug 22];78(13):1352–71. Available from: https://www.jacc.org/doi/abs/10.1016/j.jacc.2021.07.042

8. Kelshiker MA, Seligman H, Howard JP, et al. Coronary flow reserve and cardiovascular outcomes: a systematic review and meta-analysis. European Heart Journal [Internet] 2022 [cited 2022 Apr 29];43(16):1582–93. Available from: https://doi.org/10.1093/eurheartj/ehab775

9. Schindler TH, Dilsizian V. Coronary microvascular dysfunction: Clinical considerations and noninvasive diagnosis. JACC: Cardiovascular Imaging 2020;13(1):140–55.

10. Gdowski MA, Murthy VL, Doering M, Monroy-Gonzalez AG, Slart R, Brown DL. Association of isolated coronary microvascular dysfunction with mortality and major adverse cardiac events: A systematic review and meta-analysis of aggregate data. Journal of the American Heart Association [Internet] 2020 [cited 2022 May 24];9(9):e014954. Available from: https://www.ahajournals.org/doi/full/10.1161/JAHA.119.014954

11. Sanghani R, Al-Mallah MH, Thompson R. Challenges and strategies to enable access to cardiac positron emission tomography in different parts of the world: The North American perspective. Journal of Nuclear Cardiology [Internet] 2024 [cited 2024 Apr 2];31:101790. Available from: https://www.sciencedirect.com/science/article/pii/S1071358123081199

12. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nat Biomed Eng [Internet] 2022 [cited 2024 Oct 3];6(12):1346–52. Available from: https://www-nature-com.proxy.lib.umich.edu/articles/s41551-022-00914-1

13. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models [Internet]. 2022 [cited 2022 Nov 1];Available from: http://arxiv.org/abs/2108.07258

14. Zhou Y, Chia MA, Wagner SK, et al. A foundation model for generalizable disease detection from retinal images. Nature [Internet] 2023 [cited 2024 Oct 18];622(7981):156–63. Available from: https://www.nature.com/articles/s41586-023-06555-x

15. Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res [Internet] 2020;21(1):140:5485–551. Available from: https://www.jmlr.org/papers/v21/20-074.html

16. Bao H, Dong L, Piao S, Wei F. BEiT: BERT pre-training of image transformers [Internet]. 2022 [cited 2022 Nov 1];Available from: http://arxiv.org/abs/2106.08254

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

17. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [Internet]. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2017 [cited 2022 Jul 1]. Available from: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

18. Chowdhery A, Narang S, Devlin J, et al. PaLM: scaling language modeling with pathways. Journal of Machine Learning Research [Internet] 2023 [cited 2024 Mar 1];24(240):1–113. Available from: http://jmlr.org/papers/v24/22-1144.html

19. Natarajan A, Chang Y, Mariani S, et al. A wide and deep transformer neural network for 12-lead ECG classification [Internet]. 2020 [cited 2022 Jul 1]. Available from: http://www.cinc.org/archives/2020/pdf/CinC2020-107.pdf

20. Vaid A, Jiang J, Sawant A, et al. A foundational vision transformer improves diagnostic performance for electrocardiograms. npj Digit Med [Internet] 2023 [cited 2023 Jun 15];6(1):1–8. Available from: https://www.nature.com/articles/s41746-023-00840-9

21. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale [Internet]. 2021 [cited 2022 Jul 6];Available from: http://arxiv.org/abs/2010.11929

22. Gow B, Pollard T, Nathanson LA, et al. MIMIC-IV-ECG: diagnostic electrocardiogram matched subset (1.0) [Internet]. PhysioNet; 2023 [cited 2024 Feb 23]. Available from: https://physionet.org/content/mimic-iv-ecg/1.0/

23. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data [Internet] 2023 [cited 2024 Feb 23];10(1):1. Available from: https://www.nature.com/articles/s41597-022-01899-x

24. Wagner P, Strodthoff N, Bousseljot R-D, et al. PTB-XL, a large publicly available electrocardiography dataset. Sci Data [Internet] 2020 [cited 2024 Jan 2];7(1):154. Available from: https://www.nature.com/articles/s41597-020-0495-6

25. Wagner P, Strodthoff N, Bousseljot R-D, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset [Internet]. 2022 [cited 2024 Feb 23];Available from: https://physionet.org/content/ptb-xl/1.0.3/

26. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.

27. Dilsizian V, Bacharach SL, Beanlands RS, et al. ASNC imaging guidelines/SNMMI procedure standard for positron emission tomography (PET) nuclear cardiology procedures. J Nucl Cardiol [Internet] 2016 [cited 2016 Jul 18];1–40. Available from: http://link.springer.com/article/10.1007/s12350-016-0522-3

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

28. Murthy VL, Bateman TM, Beanlands RS, et al. Clinical quantification of myocardial blood flow using PET: Joint position paper of the SNMMI Cardiovascular Council and the ASNC. J Nucl Cardiol [Internet] 2018 [cited 2018 Jan 4];25:269–97. Available from: https://link.springer.com/article/10.1007/s12350-017-1110-x

29. Bateman TM, Heller GV, Beanlands R, et al. Practical guide for interpreting and reporting cardiac PET measurements of myocardial blood flow: an Information Statement from the American Society of Nuclear Cardiology, and the Society of Nuclear Medicine and Molecular Imaging. J Nucl Cardiol [Internet] 2021 [cited 2022 Feb 1];28(2):768–87. Available from: https://doi.org/10.1007/s12350-021-02552-7

30. Arida-Moody L, Moody JB, Renaud JM, et al. Effects of two patient-specific dosing protocols on measurement of myocardial blood flow with 3D ^{82}Rb cardiac PET. Eur J Nucl Med Mol Imaging 2021;48:3835–46.

31. Dorbala S, Ananthasubramaniam K, Armstrong IS, et al. Single photon emission computed tomography (SPECT) myocardial perfusion imaging guidelines: Instrumentation, acquisition, processing, and interpretation. Journal of Nuclear Cardiology [Internet] 2018 [cited 2018 Nov 14];25(5):1784–846. Available from: http://link.springer.com/10.1007/s12350-018-1283-y

32. Hagio T, Poitrasson-Rivière A, Moody JB, et al. “Virtual” attenuation correction: improving stress myocardial perfusion SPECT imaging using deep learning. European

Journal of Nuclear Medicine and Molecular Imaging [Internet] 2022;49:3140–9. Available from: https://link.springer.com/epdf/10.1007/s00259-022-05735-7

33. Ficaro EP, Lee BC, Kritzman JN, Corbett JR. Corridor4DM: the Michigan method for quantitative nuclear cardiology. J Nucl Cardiol 2007;14(4):455–65.

34. Slomka PJ, Nishina H, Berman DS, et al. Automated quantification of myocardial perfusion SPECT using simplified normal limits. J Nucl Cardiol [Internet] 2005;12(1):66–77. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1071358104004921

35. Lortie M, Beanlands RSB, Yoshinaga K, Klein R, Dasilva JN, DeKemp RA. Quantification of myocardial blood flow with ^{82}Rb dynamic PET imaging. Eur J Nucl Med Mol Imaging 2007;34(11):1765–74.

36. Moody JB, Murthy VL, Lee BC, Corbett JR, Ficaro EP. Variance estimation for myocardial blood flow by dynamic PET. IEEE Transactions on Medical Imaging 2015;34(11):2343–53.

37. Renaud JM, Poitrasson-Rivière A, Moody JB, Hagio T, Ficaro EP, Murthy VL. Improved diagnostic accuracy for coronary artery disease detection with quantitative 3D ^{82}Rb PET myocardial perfusion imaging. Eur J Nucl Med Mol Imaging [Internet] 2023 [cited 2023 Oct 3];Available from: https://doi.org/10.1007/s00259-023-06414-x

38. Dorbala S, Vangala D, Sampson U, Limaye A, Kwong R, Di Carli MF. Value of vasodilator left ventricular ejection fraction reserve in evaluating the magnitude of myocardium at risk and the extent of angiographic coronary artery disease: a ^{82}Rb PET/CT study. J Nucl Med [Internet] 2007 [cited 2008 Jun 11];48(3):349–58. Available from: http://jnm.snmjournals.org/cgi/content/abstract/48/3/349

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

39. Petersen SE, Matthews PM, Francis JM, et al. UK Biobank's cardiovascular magnetic resonance protocol. Journal of Cardiovascular Magnetic Resonance [Internet] 2016 [cited 2024 Feb 20];18(1):8. Available from: https://www.sciencedirect.com/science/article/pii/S1097664723009493

40. Bai W, Suzuki H, Huang J, et al. A population-based phenome-wide association study of cardiac and aortic structure and function. Nat Med [Internet] 2020 [cited 2024 Oct 23];26(10):1654–62. Available from: https://www.nature.com/articles/s41591-020-1009-y

41. Nakazato R, Berman DS, Dey D, et al. Automated quantitative Rb-82 3D PET/CT myocardial perfusion imaging: Normal limits and correlation with invasive coronary angiography. J Nucl Cardiol [Internet] 2012 [cited 2014 Oct 28];19(2):265–76. Available from: http://link.springer.com/article/10.1007/s12350-011-9496-3

42. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med [Internet] 2019 [cited 2022 May 26];25(1):70–4. Available from: https://www.nature.com/articles/s41591-018-0240-2

43. Ba JL, Kiros JR, Hinton GE. Layer normalization [Internet]. 2016 [cited 2023 Jan 3];Available from: http://arxiv.org/abs/1607.06450

44. Hendrycks D, Gimpel K. Gaussian error linear units (GELUs) [Internet]. 2023 [cited 2024 Feb 10];Available from: http://arxiv.org/abs/1606.08415

45. Salinet JL, Luppi Silva O. Chapter 2 - ECG Signal Acquisition Systems [Internet]. In: do Vale Madeiro JP, Cortez PC, da Silva Monteiro Filho JM, Brayner ARA, editors. Developments and Applications for ECG Signal Processing. Academic Press; 2019 [cited 2022 Nov 15]. p. 29–51.Available from: https://www.sciencedirect.com/science/article/pii/B9780128140352000086

46. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems [Internet] 2019 [cited 2024 Feb 27];32. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

47. Kaiser W, Findeis M. Artifact processing during exercise testing. Journal of Electrocardiology [Internet] 1999 [cited 2024 Jul 22];32:212–9. Available from: https://www.sciencedirect.com/science/article/pii/S0022073699900833

48. Kligfield P, Gettes LS, Bailey JJ, et al. Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I: The Electrocardiogram and Its Technology A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology. Journal of the American College of Cardiology [Internet] 2007 [cited 2016 Aug 18];49(10):1109–27. Available from: http://www.sciencedirect.com/science/article/pii/S073510970700232X

NEJMAI AUTHOR ACCEPTED MANUSCRIPT

49. Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. A transformer-based framework for multivariate time series representation learning [Internet]. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York, NY, USA: Association for Computing Machinery; 2021 [cited 2022 Nov 8]. p. 2114–24.Available from: https://doi.org/10.1145/3447548.3467401

50. Xie Z, Zhang Z, Cao Y, et al. SimMIM: a simple framework for masked image modeling [Internet]. 2022 [cited 2023 Feb 15]. p. 9653–63.Available from: https://openaccess.thecvf.com/content/CVPR2022/html/Xie_SimMIM_A_Simple_Framework_for_Masked_Image_Modeling_CVPR_2022_paper.html

51. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners [Internet]. 2022 [cited 2022 Nov 1]. p. 16000–9.Available from: https://openaccess.thecvf.com/content/CVPR2022/html/He_Masked_Autoencoders_Are_Scalable_Vision_Learners_CVPR_2022_paper.html

52. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:14126980 [cs] [Internet] 2014 [cited 2019 Jul 9];Available from: http://arxiv.org/abs/1412.6980

53. Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts [Internet]. 2017 [cited 2022 Nov 17];Available from: http://arxiv.org/abs/1608.03983

54. Steiner A, Kolesnikov A, Zhai X, Wightman R, Uszkoreit J, Beyer L. How to train your ViT? Data, augmentation, and regularization in vision transformers [Internet]. 2022 [cited 2024 Feb 27];Available from: http://arxiv.org/abs/2106.10270

55. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision [Internet]. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE; 2016 [cited 2024 Jul 13]. p. 2818–26.Available from: http://ieeexplore.ieee.org/document/7780677/

56. Müller R, Kornblith S, Hinton GE. When does label smoothing help? [Internet]. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2019 [cited 2024 Jul 13]. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/hash/f1748d6b0fd9d439f71450117eba2725-Abstract.html

57. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Statistics in Medicine [Internet] 2019 [cited 2024 Apr 29];38(21):4051–65. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8281

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

58. Wang Y, Huang H, Rudin C, Shaposhnik Y. Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. Journal of Machine Learning Research [Internet] 2021 [cited 2025 Sep 19];22(201):1–73. Available from: http://jmlr.org/papers/v22/20-1061.html

59. Wang Y, Sun Y, Huang H, Rudin C. Dimension reduction with locally adjusted graphs. Proceedings of the AAAI Conference on Artificial Intelligence [Internet] 2025 [cited 2025 Sep 21];39(20):21357–65. Available from: https://ojs.aaai.org/index.php/AAAI/article/view/35436

60. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics [Internet] 1988;44(3):837–45. Available from: http://www.jstor.org/stable/2531595

61. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika [Internet] 1947 [cited 2019 Aug 12];12(2):153–7. Available from: https://doi.org/10.1007/BF02295996

62. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. Biometrics [Internet] 2000 [cited 2019 Aug 12];56(2):345–51. Available from: https://www.jstor.org/stable/2676972

63. Strodthoff N, Wagner P, Schaeffter T, Samek W. Deep learning for ECG analysis: benchmarks and insights from PTB-XL. IEEE Journal of Biomedical and Health Informatics [Internet] 2021 [cited 2023 Dec 21];25(5):1519–28. Available from: https://ieeexplore.ieee.org/document/9190034

64. Bjerkén LV, Rønborg SN, Jensen MT, Ørting SN, Nielsen OW. Artificial intelligence enabled ECG screening for left ventricular systolic dysfunction: a systematic review. Heart Fail Rev [Internet] 2023 [cited 2024 May 16];28(2):419–30. Available from: https://doi.org/10.1007/s10741-022-10283-1

65. Starc V, Leban MA, Sinigoj P, et al. Can functional cardiac age be predicted from the ECG in a normal healthy population? In: 2012 Computing in Cardiology. Krakow, Poland: IEEE; 2012. p. 101–4.

66. Ball RL, Feiveson AH, Schlegel TT, Starc V, Dabney AR. Predicting "Heart Age" using electrocardiography. Journal of Personalized Medicine [Internet] 2014 [cited 2024 Oct 14];4(1):65–78. Available from: https://www.mdpi.com/2075-4426/4/1/65

67. Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. Circulation: Arrhythmia and Electrophysiology [Internet] 2019 [cited 2023 Aug 26];12(9):e007284. Available from: https://www.ahajournals.org/doi/full/10.1161/CIRCEP.119.007284

68. Pastika L, Sau A, Patlatzoglou K, et al. Artificial intelligence-enhanced electrocardiography derived body mass index as a predictor of future cardiometabolic disease. npj Digit Med [Internet] 2024 [cited 2024 Sep 10];7(1):1–16. Available from: https://www-nature-com.proxy.lib.umich.edu/articles/s41746-024-01170-0

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

69. Ansari MY, Qaraqe M, Charafeddine F, Serpedin E, Righetti R, Qaraqe K. Estimating age and gender from electrocardiogram signals: A comprehensive review of the past decade. Artificial Intelligence in Medicine [Internet] 2023 [cited 2024 Jul 5];146:102690. Available from: https://www.sciencedirect.com/science/article/pii/S093336572300204X

70. Moody JB, Poitrasson-Rivière A, Hagio T, et al. Added value of myocardial blood flow using ^{18}F-flurpiridaz PET to diagnose coronary artery disease: The flurpiridaz 301 trial. Journal of Nuclear Cardiology [Internet] 2021;28(5):2313–29. Available from: https://pubmed.ncbi.nlm.nih.gov/32002847/

71. Murthy VL, Naya M, Foster CR, et al. Improved cardiac risk assessment with noninvasive measures of coronary flow reserve. Circulation [Internet] 2011 [cited 2012 May 17];124(20):2215–24. Available from: http://circ.ahajournals.org/content/124/20/2215

72. Tamarappoo BK, Ling LF, Cerqueira M, Hachamovitch R. Independent prognostic value of left ventricular contractile reserve and chronotropic response in patients with reduced left ventricular ejection fraction undergoing vasodilator stress myocardial perfusion imaging with Rb-82 positron emission tomography. Eur Heart J Cardiovasc

Imaging [Internet] 2018;19(4):442–9. Available from: https://academic.oup.com/ehjcimaging/article/19/4/442/3902995

73. Ericsson L, Gouk H, Loy CC, Hospedales TM. Self-supervised representation learning: introduction, advances, and challenges. IEEE Signal Processing Magazine [Internet] 2022 [cited 2024 Apr 18];39(3):42–62. Available from: https://ieeexplore.ieee.org/abstract/document/9770283

74. Ahmad A, Shelly-Cohen M, Corban MT, et al. Machine learning aids clinical decision-making in patients presenting with angina and non-obstructive coronary artery disease. European Heart Journal - Digital Health [Internet] 2021 [cited 2023 Aug 26];2(4):597–605. Available from: https://doi.org/10.1093/ehjdh/ztab084

75. Alahdab F, Saad MB, Ahmed AI, et al. Development and validation of a machine learning model to predict myocardial blood flow and clinical outcomes from patients' electrocardiograms. Cell Reports Medicine [Internet] 2024 [cited 2024 Sep 26];101746. Available from: https://www.sciencedirect.com/science/article/pii/S2666379124004762

76. Li J, Aguirre AD, Junior VM, et al. An electrocardiogram foundation model built on over 10 million recordings. NEJM AI [Internet] 2025 [cited 2025 Jun 29];2(7):AIoa2401033. Available from: https://ai.nejm.org/doi/10.1056/AIoa2401033

77. McKeen K, Oliva L, Masood S, Toma A, Rubin B, Wang B. ECG-FM: An open electrocardiogram foundation model [Internet]. 2024 [cited 2024 Nov 12];Available from: http://arxiv.org/abs/2408.05178

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

78. Nolin-Lapalme A, Sowa A, Delfrate J, et al. Foundation models for generalizable electrocardiogram interpretation: comparison of supervised and self-supervised electrocardiogram foundation models [Internet]. 2025 [cited 2025 Mar 10];2025.03.02.25322575. Available from: https://www.medrxiv.org/content/10.1101/2025.03.02.25322575v1

79. Sangha V, Mortazavi BJ, Haimovich AD, et al. Automated multilabel diagnosis on electrocardiographic images and signals. Nat Commun [Internet] 2022 [cited 2024 Oct 9];13(1):1583. Available from: https://www.nature.com/articles/s41467-022-29153-3

80. Duong SQ, Vaid A, My VTH, et al. Quantitative prediction of right ventricular size and function from the ECG. Journal of the American Heart Association [Internet] 2024 [cited 2024 Jan 25];13(1):e031671. Available from: https://www.ahajournals.org/doi/10.1161/JAHA.123.031671

81. Sara JD, Widmer RJ, Matsuzawa Y, Lennon RJ, Lerman LO, Lerman A. Prevalence of coronary microvascular dysfunction among patients with chest pain and nonobstructive coronary artery disease. J Am Coll Cardiol Intv [Internet] 2015 [cited 2019 Oct 11];8(11):1445–53. Available from: http://interventions.onlinejacc.org/content/8/11/1445

82. Masi S, Rizzoni D, Taddei S, et al. Assessment and pathophysiology of microvascular disease: recent progress and clinical implications. European Heart Journal [Internet] 2021 [cited 2023 Aug 23];42(26):2590–604. Available from: https://doi.org/10.1093/eurheartj/ehaa857

83. Jespersen L, Hvelplund A, Abildstrøm SZ, et al. Stable angina pectoris with no obstructive coronary artery disease is associated with increased risks of major adverse cardiovascular events. European Heart Journal [Internet] 2012 [cited 2022 Feb 9];33(6):734–44. Available from: https://doi.org/10.1093/eurheartj/ehr331

84. Ong P, Safdar B, Seitz A, Hubert A, Beltrame JF, Prescott E. Diagnosis of coronary microvascular dysfunction in the clinic. Cardiovascular Research [Internet] 2020 [cited 2023 Aug 23];116(4):841–55. Available from: https://doi.org/10.1093/cvr/cvz339

85. Feher A, Sinusas AJ. Quantitative assessment of coronary microvascular function: dynamic single-photon emission computed tomography, positron emission tomography, ultrasound, computed tomography, and magnetic resonance imaging. Circulation: Cardiovascular Imaging [Internet] 2017 [cited 2018 Aug 23];10(8):e006427. Available from: http://circimaging.ahajournals.org/lookup/doi/10.1161/CIRCIMAGING.117.006427

86. Arbelo E, Protonotarios A, Gimeno JR, et al. 2023 ESC Guidelines for the management of cardiomyopathies: Developed by the task force on the management of cardiomyopathies of the European Society of Cardiology (ESC). European Heart Journal [Internet] 2023 [cited 2024 May 9];44(37):3503–626. Available from: https://doi.org/10.1093/eurheartj/ehad194

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

87. Delgado V, Ajmone Marsan N, de Waha S, et al. 2023 ESC Guidelines for the management of endocarditis: Developed by the task force on the management of endocarditis of the European Society of Cardiology (ESC) Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS) and the European Association of Nuclear Medicine (EANM). European Heart Journal [Internet] 2023 [cited 2024 May 9];44(39):3948–4042. Available from: https://doi.org/10.1093/eurheartj/ehad193

88. Virani SS, Newby LK, Arnold SV, et al. 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA Guideline for the Management of Patients With Chronic Coronary Disease. Journal of the American College of Cardiology [Internet] 2023 [cited 2024 May 9];82(9):833–955. Available from: https://www.jacc.org/doi/10.1016/j.jacc.2023.04.003

89. Lima RSL, Watson DD, Goode AR, et al. Incremental value of combined perfusion and function over perfusion alone by gated SPECT myocardial perfusion imaging for detection of severe three-vessel coronary artery disease. Journal of the American College of Cardiology [Internet] 2003 [cited 2019 May 28];42(1):64–70. Available from: http://www.onlinejacc.org/content/42/1/64

90. Gould KL, Johnson NP, Bateman TM, et al. Anatomic versus physiologic assessment of coronary artery disease. Role of coronary flow reserve, fractional flow reserve, and positron emission tomography imaging in revascularization decision-making. J Am Coll

Cardiol [Internet] 2013 [cited 2014 Aug 26];62(18):1639–53. Available from: http://dx.doi.org/10.1016/j.jacc.2013.07.076

91. Patel KK, Singh A, Qarajeh R, et al. Prevalence of balanced ischemia on PET relative perfusion imaging among patients with high-risk coronary artery disease. Journal of Nuclear Cardiology [Internet] 2024 [cited 2024 Mar 8];101836. Available from: https://www.sciencedirect.com/science/article/pii/S1071358124004549

92. Kajander S, Joutsiniemi E, Saraste M, et al. Clinical value of absolute quantification of myocardial perfusion with ^{15}O-water in coronary artery disease. Circulation: Cardiovascular Imaging [Internet] 2011 [cited 2019 Aug 12];4(6):678–84. Available from: insights.ovid.com

93. Ziadi MC, deKemp RA, Williams K, et al. Does quantification of myocardial flow reserve using rubidium-82 positron emission tomography facilitate detection of multivessel coronary artery disease? J Nucl Cardiol [Internet] 2012 [cited 2014 Nov 24];19(4):670–80. Available from: http://link.springer.com/article/10.1007/s12350-011-9506-5

94. Poitrasson-Rivière A, Moody JB, Renaud JM, et al. Integrated myocardial flow reserve (iMFR) assessment: optimized PET blood flow quantification for diagnosis of coronary artery disease. European Journal of Nuclear Medicine and Molecular Imaging 2023;51:136–46.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

95. Moody JB, Poitrasson-Rivière A, Renaud JM, et al. Integrated myocardial flow reserve (iMFR) assessment: diffuse atherosclerosis and microvascular dysfunction are more strongly associated with mortality than focally impaired perfusion. Eur J Nucl Med Mol Imaging [Internet] 2023 [cited 2023 Oct 3];51:123–35. Available from: https://doi.org/10.1007/s00259-023-06448-1

Tables

Table 1: ECG databases used in this study.

ECG Database	Source	Imaging Modality	Subjects	ECGs
MIMIC-IV-ECG	physionet.org	--	161,352	800,035
UM-PET	University of Michigan	PET/CT	5,198	10,396
HM-PET	Houston-Methodist Hospital	PET/CT	758	1,516
UM-phSPECT	University of Michigan	SPECT/CT	5,102	10,204
UM-exSPECT	University of Michigan	SPECT/CT	1,533	3,066
PTB-XL	physionet.org	--	18,969	21,799
UK-Biobank	ukbiobank.ac.uk	CMR	66,670	66,670
Total			259,582	913,686

All PET/CT and SPECT/CT databases included both a rest and stress ECG from individual imaging exams for each patient.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Table 2: Prediction tasks and ECG databases used for pretraining, fine-tuning and testing the SSL foundation model (P=SSL pretraining, F=Fine-tuning, H=Holdout testing, E=External testing, O=Internal, cross-modality testing).

Task	MIMIC-IV-ECG	UM-PET	HM-PET	UM-phSPECT	UM-exSPECT	PTB-XL	UK-Biobank
LV Myocardial Perfusion							
MFR	P	F/H	E	—	—	—	—
Rest MBF	P	F/H	E	—	—	—	—
Stress MBF	P	F/H	E	—	—	—	—
Stress TPD	P	F/H	E	O	O	—	—
LV Myocardial Function							
LVEF	P	F/H	E	O	O	—	E
LVEF Reserve	P	F/H	E	—	—	—	—
Auto ECG Interpretation							
Diagnostic	P	—	—	—	—	F/H	—
Rhythm	P	—	—	—	—	F/H	—
Form	P	—	—	—	—	F/H	—
Demographics							
Age	P	F/H	E	O	O	E	E
BMI	P	F/H	E	O	O	E	E
Sex	P	F/H	E	O	O	E	E

MIMIC-IV-ECG data was unlabeled and used for self-supervised pretraining of foundation model.

Labels for LV myocardial perfusion and function tasks were derived from PET imaging.

MFR=myocardial flow reserve (threshold < 2).

Rest MBF=rest myocardial blood flow (threshold > 1.2 ml/min/g).

Stress MBF threshold < 1.5 ml/min/g.

Stress TPD=stress total perfusion deficit (threshold > 5%).

LVEF=left ventricular ejection fraction (threshold < 35%).

LVEF Reserve threshold < -5%.

Auto ECG Interpretation labels derived from clinical ECG reports.

Demographics labels derived from patient records.

BMI=body mass index.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Table 3: Baseline characteristics of patients who underwent stress testing by myocardial perfusion imaging in the UM-PET and UM-phSPECT databases.

	UM-PET		UM-phSPECT	
	Model Derivation	Holdout Test	Internal Test	p-value
n	4167	1031	5102	
Age (y)	63 (12)	63 (12)	66 (12)	<0.001
Male Sex	2312 (55)	590 (57)	2874 (56)	<0.001
White	3404 (82)	840 (81)	4127 (81)	<0.001
Black	512 (12)	115 (11)	504 (10)	
Other	251 (6)	76 (7)	471 (9)	
Body mass index (kg/m^2)	34 (10)	33 (9)	30 (6)	<0.001
Obesity	2539 (61)	599 (58)	2303 (45)	<0.001
Hypertension	3179 (76)	776 (75)	4017 (79)	<0.001
Diabetes	1652 (40)	422 (41)	1676 (33)	<0.001
Hyperlipidemia	2747 (66)	673 (65)	3699 (73)	<0.001
Family history of coronary artery disease	2807 (67)	681 (66)	1715 (34)	<0.001
Current smoker	336 (8)	99 (10)	1135 (22)	<0.001
History of myocardial infarction	799 (19)	200 (19)	714 (14)	<0.001
Prior percutaneous coronary intervention	816 (20)	201 (19)	952 (19)	<0.001
Prior coronary artery bypass graft	418 (10)	102 (10)	491 (10)	<0.001
Angina	1819 (44)	431 (42)	2204 (43)	<0.001
Dyspnea	665 (16)	172 (17)	1077 (21)	<0.001
Heart rate, rest (bpm)	73 (14)	73 (13)	69 (13)	<0.001
SBP, rest (mm Hg)	132 (23)	131 (24)	143 (25)	<0.001
DBP, rest (mm Hg)	71 (24)	71 (22)	80 (13)	<0.001
Rate pressure product (mm Hg bpm)	9658 (2496)	9527 (2470)	9853 (2522)	<0.001
Heart rate, stress (bpm)	93 (20)	92 (16)	101 (20)	<0.001
SBP, stress (mm Hg)	120 (31)	119 (30)	161 (27)	<0.001
DBP, stress (mm Hg)	64 (39)	63 (36)	80 (14)	<0.001

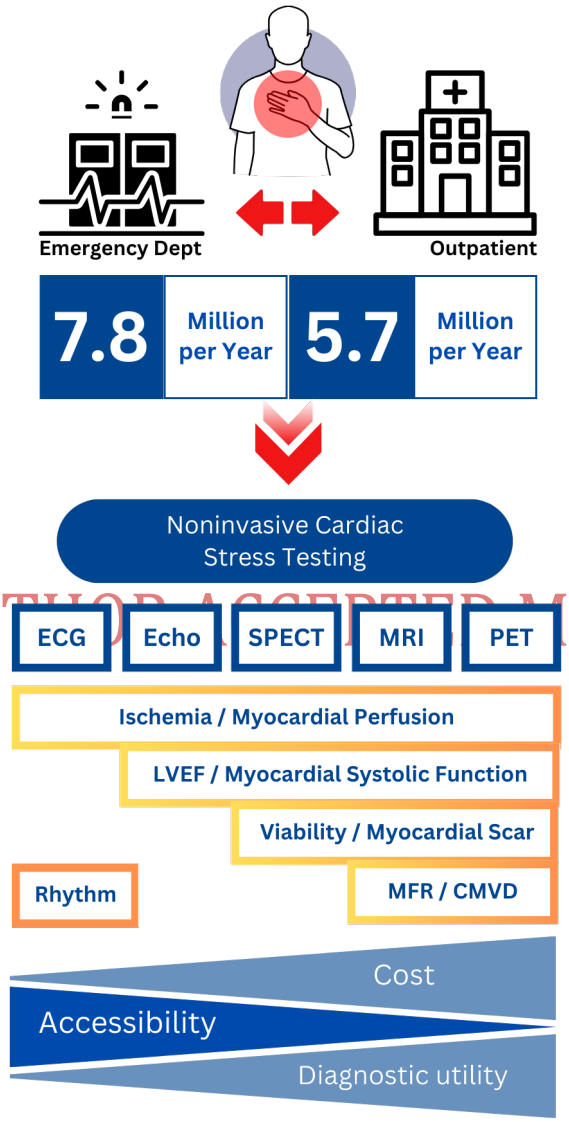
NEJM AI AUTHOR ACCEPTED MANUSCRIPT

LV ejection fraction, rest (%)	57 (15)	58 (15)	—	0.806
LV ejection fraction, stress (%)	61 (16)	61 (16)	65 (13)	<0.001
Total perfusion deficit, stress (%)	7 (10)	7 (11)	5 (9)	<0.001
Myocardial blood flow, rest (ml/min/g)	1.08 (0.39)	1.08 (0.41)	—	0.929
Myocardial blood flow, stress (ml/min/g)	2.17 (0.83)	2.16 (0.82)	—	0.784
Myocardial flow reserve	2.11 (0.72)	2.12 (0.74)	—	0.884

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Figures

Figure 1: Chest pain encounters in the United States and available noninvasive testing modalities. From left to right, as modality cost increases and accessibility decreases, the traditional capabilities and diagnostic utility for assessing cardiac and coronary function tend to increase.



NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Figure 2: External tests evaluated in the HM-PET database. Task performance was compared with and without SSL pretraining. The SSL pretrained model was fine-tuned in the UM-PET model derivation cohort. The non-SSL model underwent supervised de novo training from random initialization in the same UM-PET cohort. A complete summary is given in Supplementary Table S2.

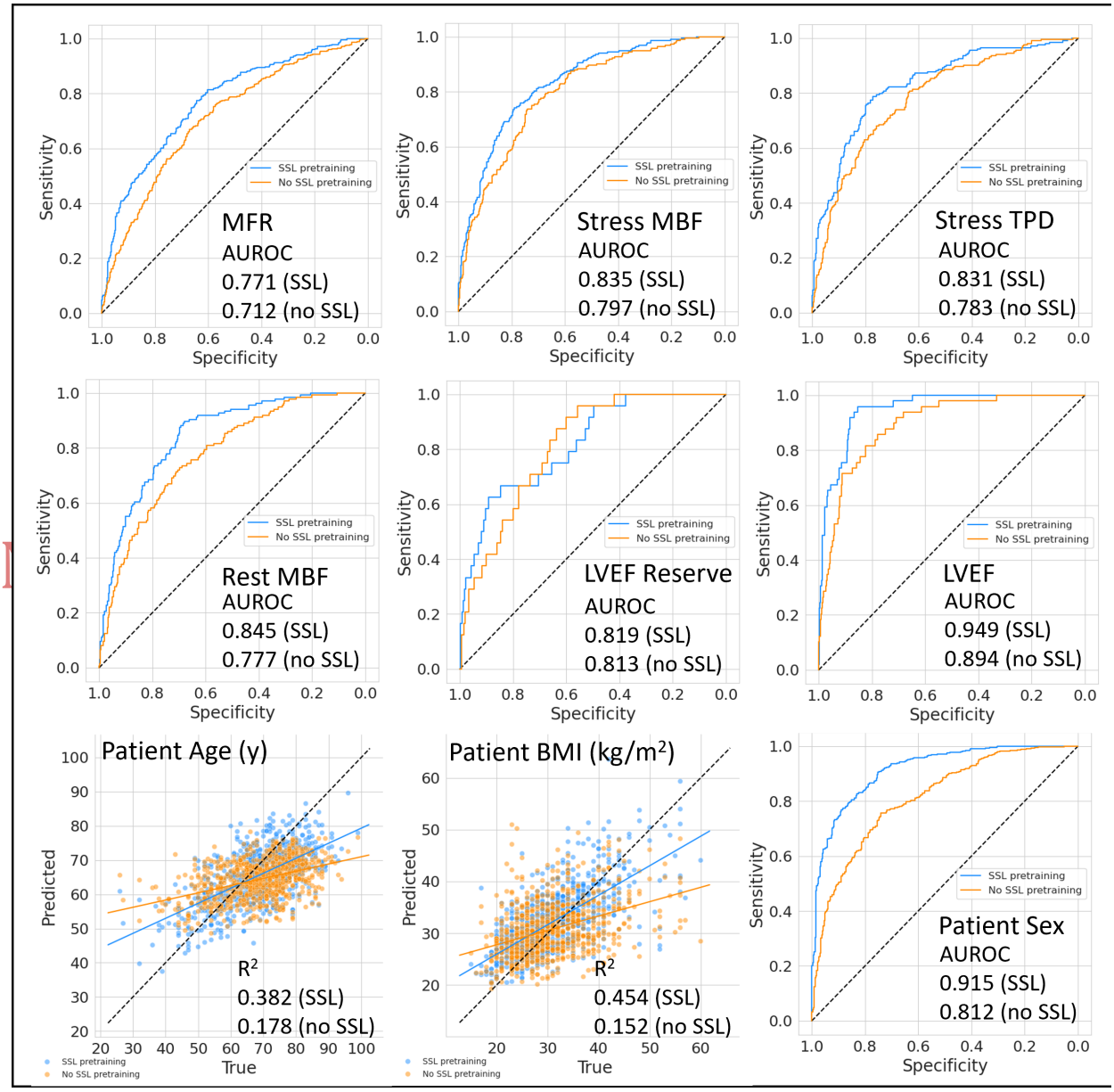


Figure 3: Additional tests of the SSL pretrained model in the UM-phSPECT, UM-exSPECT, UK Biobank, and PTB-XL databases.

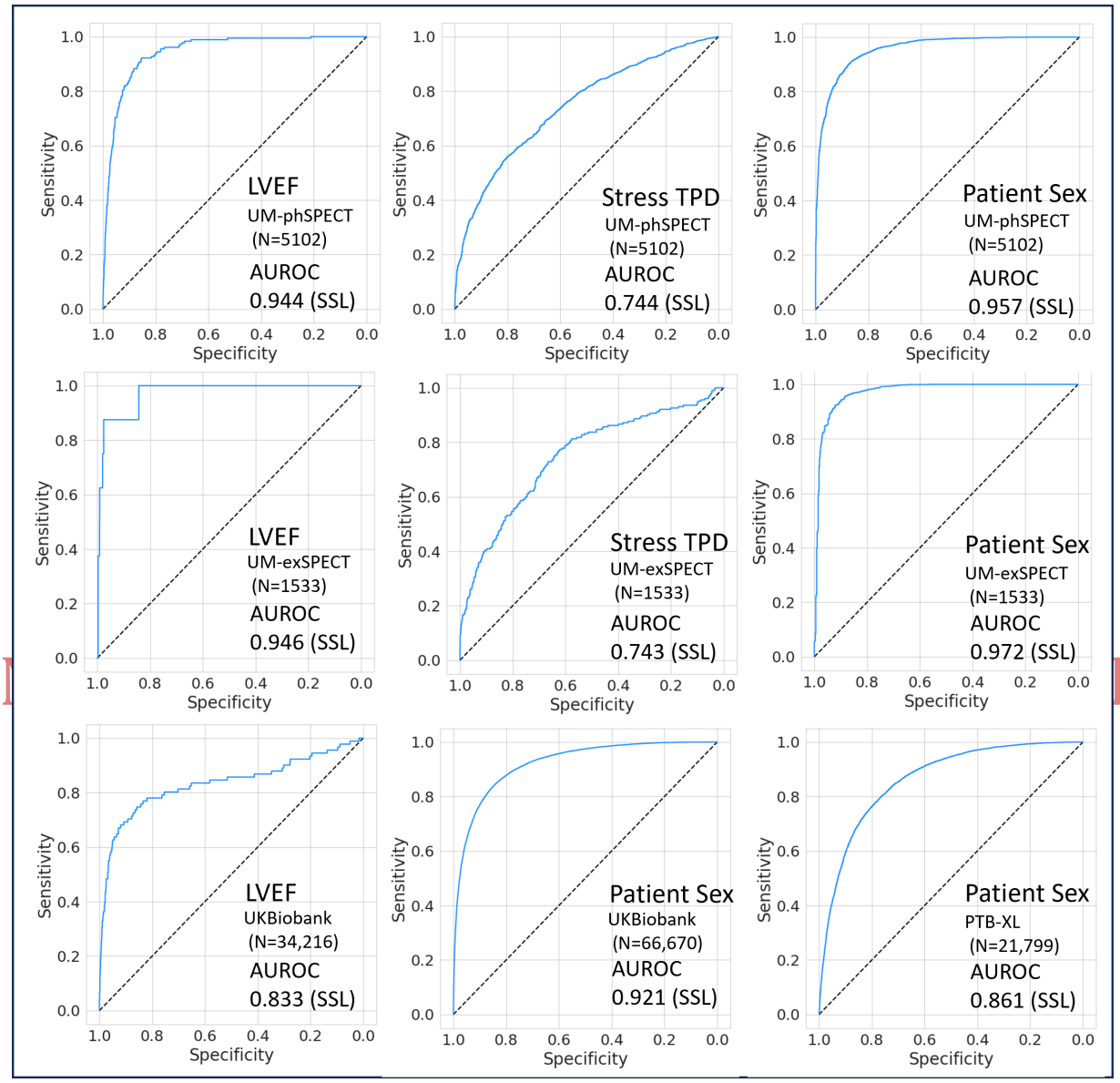
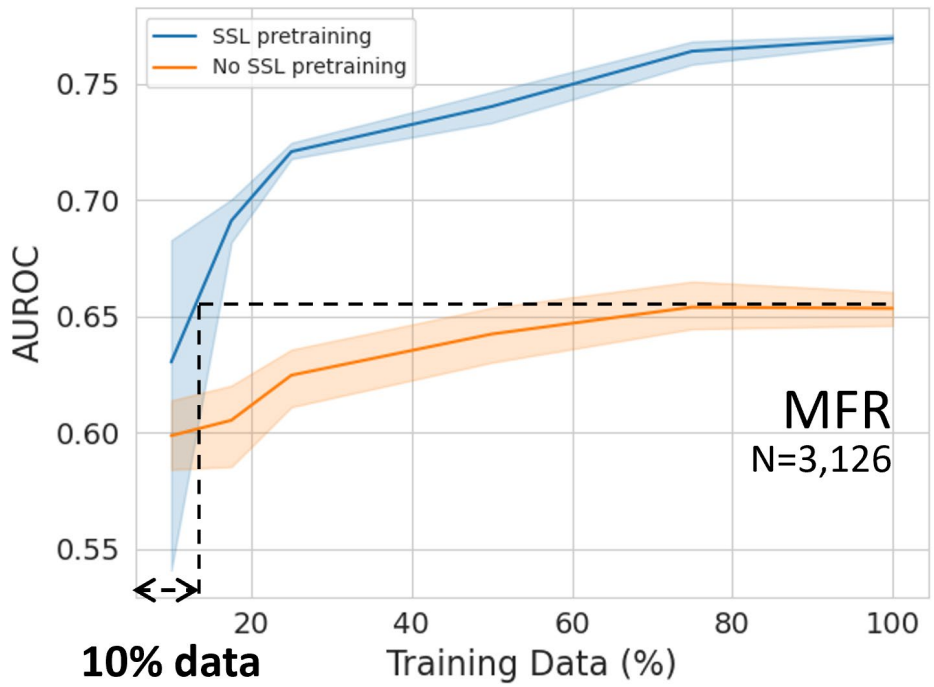


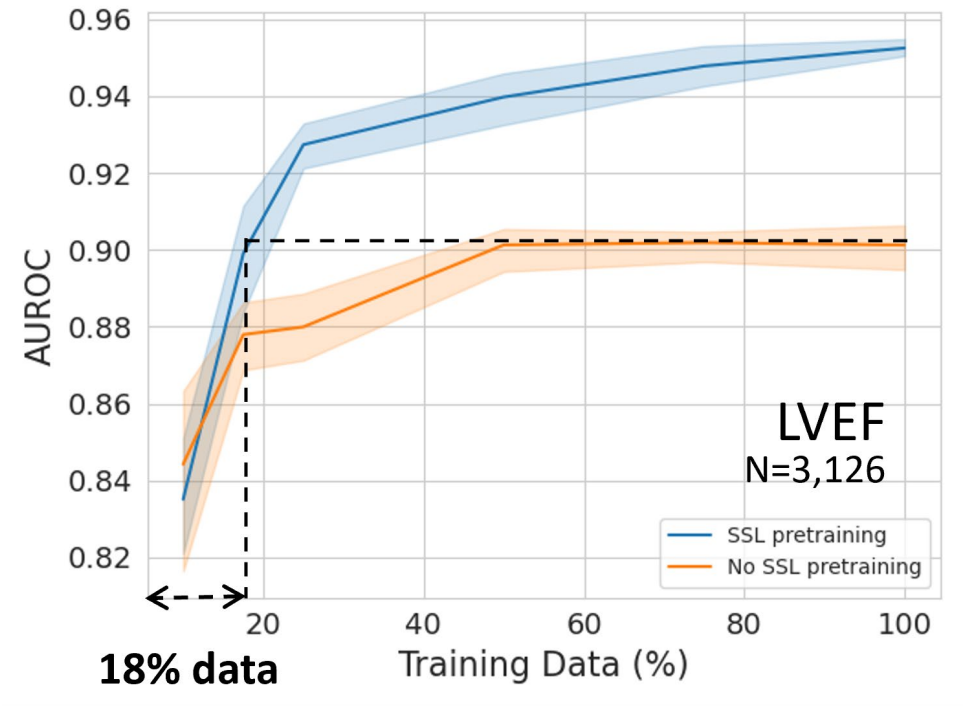
Figure 4: Evaluation of label efficiency. AUROC versus the fraction of training data used for supervised fine-tuning after SSL pretraining and de novo training with no SSL pretraining. (a) MFR task and (b) LVEF task in the UM-PET holdout database; (c) auto ECG diagnosis task and (d) auto ECG rhythm task in the PTB-XL database.

(a)



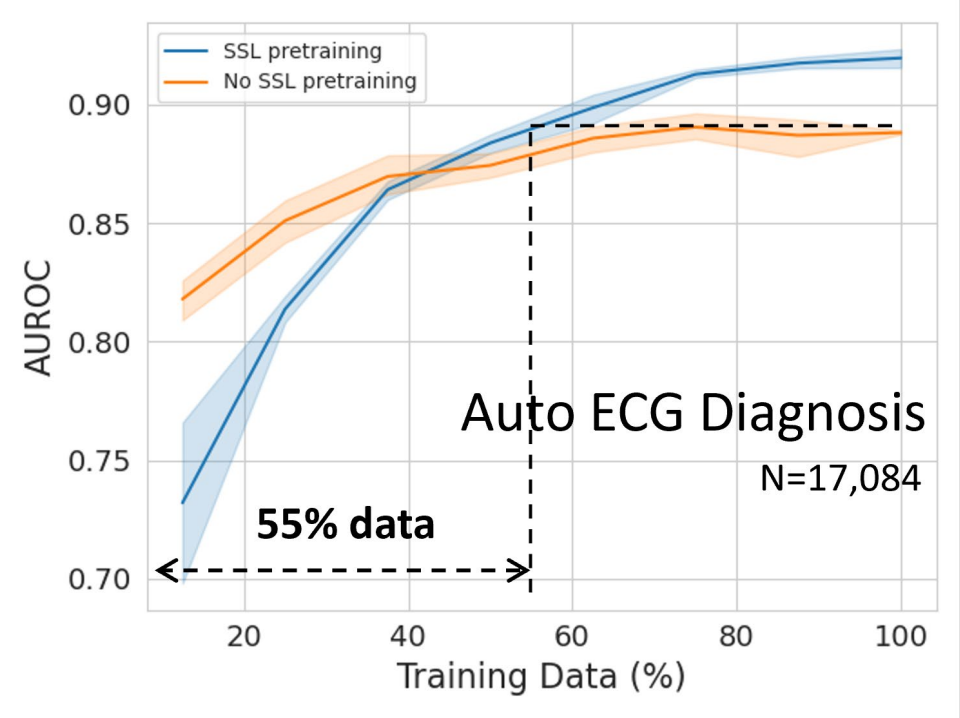
N USCRIPT

(b)

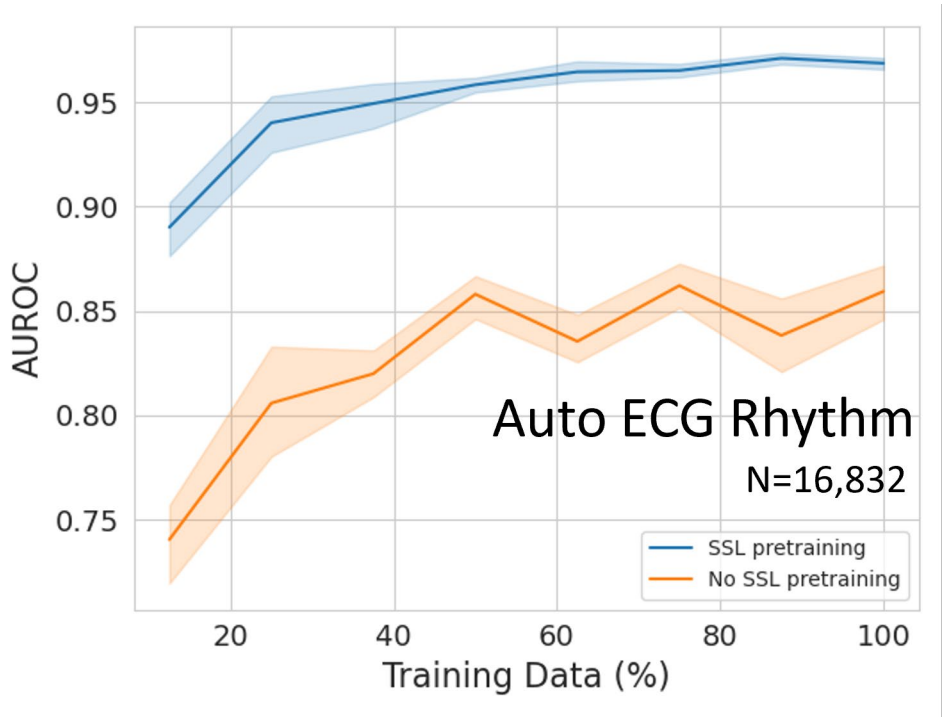


NEJM AI AUTHOR ACCEPTED MANUSCRIPT

(c)



(d)



NEJM AI AUTHOR ACCEPTED MANUSCRIPT