CARDIOPRIME: CARDIOVASCULAR PHYSIOLOGICAL REPRESENTATION INTEGRATION WITH MULTIMODAL EMBEDDINGS

Zachary Levine
Faculty of Mathematics and Computer Science
Weizmann Institute of Science
Rehovot, Israel
{zachary.levine}@weizmann.ac.il

Hagai Rossman
Pheno.Ai
Tel Aviv, Israel
{hagai}@pheno.ai

Eran Segal
Faculty of Mathematics and Computer Science
Weizmann Institute of Science
Rehovot, Israel
{eran.segal}@weizmann.ac.il

ABSTRACT

We introduce **CardioPRIME**, a hybrid mechanistic–deep learning framework for electrocardiogram (ECG) analysis, which integrates clinical phenotypes alongside classical ordinary differential equation-based cardiac models. Despite the efficacy of purely data-driven neural networks, real-world medical data are rooted in decades of physiological research. By enforcing consistency between ECG-derived latent representations and multiple clinical modalities, CardioPRIME produces more discriminative and physiologically grounded embeddings than a non-integrated baseline. Our results, reflected by higher clustering metrics (NMI, AMI, Homogeneity, and Completeness) across top prevalent diseases, underscore the significance of clinical integration for improved disease separation and enhanced interpretability. This indicates that bridging physiological modeling with data-driven techniques can substantially advance representation learning in the cardiovascular domain for general disease diagnosis.

INTRODUCTION

RELATED WORK

Electrocardiograms (ECGs) are a fundamental tool in cardiovascular diagnostics. Recent advances in deep learning have demonstrated strong performance in electrocardiogram (ECG) analysis Diamant et al. (2021); Kiyasseh et al. (2021); McKeen et al. (2024); Friedman et al. (2025). However, purely data-driven models often overlook well-established physiological insights, limiting interpretability and clinical utility.

Mechanistic–deep learning models aim to bridge this gap by embedding *physiological priors* into neural architectures Miller et al. (2021); Chen et al. (2019). Ordinary differential equation (ODE)-based models of cardiac electrophysiology McSharry et al. (2003) describe the P–QRS–T complex in terms of interpretable parameters, allowing for more structured representations Cli (2006). These approaches have shown promise for ECG generation Golany et al. (2021); Yehuda & Radinsky (2024), classification Golany et al. (2020), and representation learning van de Leur et al. (2022).

At the same time, modern clinical datasets provide **multimodal phenotyping**, including DEXA scans, continuous glucose monitors (CGM), ultrasound, and retinal imaging, which contain complementary health information Linial et al. (2021); Salvador et al. (2024). Despite their relevance, these features are rarely integrated into ECG-based representation learning.

In this paper, we propose **CardioPRIME** (*Cardiovascular Physiological Representation Integration with Multimodal Embeddings*):

1. A **hybrid generative model** that combines ODE-based cardiac modeling with varational autoencoders,
2. A strategy for **integrating multimodal phenotyping** by aligning ECG and clinical latent spaces,
3. A demonstration of **improved disease separation** through unsupervised clustering analysis.

By aligning ECG-derived representations with multimodal phenotyping data, CardioPRIME provides a physiologically meaningful latent space with enhanced clinical applicability.

METHODS

DATA COLLECTION

Our dataset is derived from The Human Phenotype Project, a large-scale longitudinal study of Israeli adults aged 25–70, and centers on **12-lead ECG time series** recorded for four seconds at 1000hz, using the 12 Lead NORAV ECG machine with an integrated electrodes chest belt for precordial leads. We paired these ECG recordings with data from each of the following tabular clinical domains j.

1. **Dual Energy X-ray Absorptiometry (DEXA) scans** (body composition, bone mineral density),
2. **Continuous Glucose Monitor (CGM)**-extracted clinical glycemic variance and control features using 'iglu' Broll et al. (2021) (glucose management index, estimated HbA1C),
3. **Fundus imaging**-derived retinal microvasculature features using AutoMorph Zhou et al. (2022) (vein vessel density of both eyes, artery width),
4. **Liver ultrasound (shear-wave elastography)** measurements (sound speed, attenuation),
5. **Anthropometric data** (height, weight, BMI).

All personal identifiers were removed before analysis per institutional IRB approvals.

ODE-BASED CARDIAC MODEL

Our model is based on three learned modules.

- **ECG Encoder**: A bidirectional LSTM maps four-second 12-lead ECG windows to a 50-dimensional latent space $\mathbf{z}_{\text{LSTM}}$.
- **Clinical MLP**: A multi-layer perceptron maps clinical features to the *same* latent space $\mathbf{z}_{\text{MLP}}$.
- **Latent Alignment**: A KL divergence term aligns ECG and clinical latent distributions to enforce consistency.

We incorporate the McSharry ODE-based cardiac model McSharry et al. (2003) to enforce physiological constraints on ECG representations. The equations model the P–QRS–T complex and maintain interpretable latent parameters Cli (2006). As illustrated in Figure 1, for each patient, we begin by passing a four-second window of their 12-lead ECG time series, denoted as $\text{ECG}_{\text{original},i}$, to a bidirectional long short-term memory (LSTM) Hochreiter & Schmidhuber (1997) encoder. The LSTM encoder, defined as $f_{\text{LSTM}}(\cdot)$, processes the ECG data in both forward and backward directions, generating hidden states that capture temporal dependencies:

$$\mathbf{h}_i = f_{\text{LSTM}}(\text{ECG}_{\text{original},i}),$$

where $\mathbf{h}_i \in \mathbb{R}^{d_h}$ represents the concatenated hidden states. These hidden states are then projected into a 50-dimensional embedding space through a linear transformation:

$$\mathbf{z}_{\text{LSTM},i} = \mathbf{W}_{\text{LSTM}}\mathbf{h}_i + \mathbf{b}_{\text{LSTM}},$$

where $\mathbf{z}_{\text{LSTM},i} \in \mathbb{R}^{50}$ represents the latent embedding for the ECG data. Our model is a variational model. After sampling, the latent embedding is reduced to a 25-dimensional feature space, denoted as $\mathbf{z}_{\text{latent},i}$, used for ECG simulation. Simultaneously, for a selected clinical modality j, a two-layer multi-layer perceptron (MLP), denoted as $f_{\text{MLP},j}(\cdot)$, processes that patient's corresponding clinical features $\text{Clinical}_{i,j}$. The MLP maps these clinical features to the same latent space as the LSTM encoder:

$$\mathbf{z}_{\text{MLP},i,j} = f_{\text{MLP},j}(\text{Clinical}_{i,j}),$$

where $\mathbf{z}_{\text{MLP},i,j} \in \mathbb{R}^{50}$. To ensure consistent representations between the LSTM and MLP networks, we incorporate a Kullback–Leibler (KL) divergence term into the loss function, aligning the posterior distributions of the latents learned from both modalities, $q_{\text{MLP},j}(\cdot)$ and $q_{\text{LSTM}}(\cdot)$. The latent parameters $\mathbf{z}_{\text{latent},i}$ from the LSTM encoder are fed into a classical Euler solver, which evolves the learned initial states forward in time using McSharry's system, generating synthetic ECG trajectories. To minimize computational overhead, we do not apply the Euler solver to Clinical MLP-extracted parameters, instead relying solely on latent space alignment. To map synthetic trajectories back to the raw ECG domain over all 12 leads, we employ a unidirectional LSTM decoder, $f_{\text{decoder}}(\cdot)$, where

$$\text{ECG}_{\text{reconstructed},i}, = f_{\text{decoder}}(\mathbf{z}_{\text{latent},i}),$$

The model is trained using the following loss function:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \|\text{ECG}_{\text{reconstructed},i} - \text{ECG}_{\text{original},i}\|_2^2 + \alpha_1 \text{KL}(q_{\text{MLP}} \| q_{\text{LSTM}}) + \alpha_2 \text{KL}(p_{\text{prior}} \| q_{\text{LSTM}})$$

where α_1 and α_2 enforce multimodal and physiological alignment.

Where the $\alpha_1 = 2, \quad \alpha_2 = 1.5, \quad \lambda_1 = 5e-2$, and $\mathbf{J}\|_F^2$ is the Frobenius Norm of the Jacobian of the system, which helps with stability Finlay et al. (2020).

In keeping with previous work Wang & Fox (2023), after sampling, we constrained all parameters to the ranges in Table 1 using a weighted combination of the sigmoid, hyperbolic tangent, and natural logarithm functions, with learned weights initialized from a uniform distribution, and constrained to a valid distribution using the softmax function.

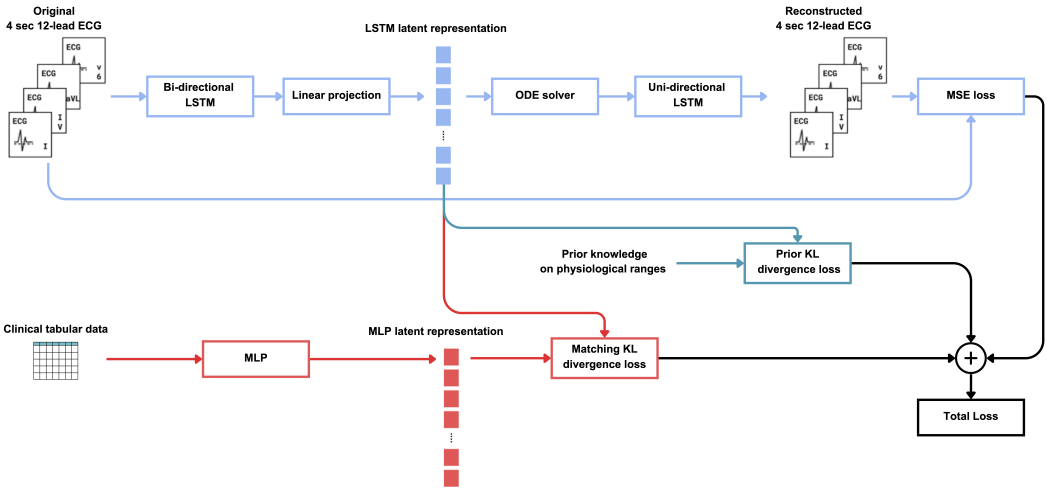

Figure 1: **CardioPRIME Model Architecture for Hybrid Mechanistic–Deep Learning.** A four-second, 12-lead ECG trace (left) is processed by a bidirectional LSTM, which projects the time-series into a 50-dimensional latent space. We then downsample to a 25-dimensional representation, which is evolved forward in time by a classical ODE solver (McSharry's model), thereby encoding explicit physiological constraints in the latent dynamics. Separately, clinical tabular data (such as DEXA, CGM, or Ultrasound measurements) are passed through an MLP to produce an aligned latent representation. A KL divergence term (*Matching KL divergence loss*) forces the ECG-based and clinical-based latents to agree, while a second KL term (*Prior KL divergence loss*) keeps the parameters within physiological norms. A unidirectional LSTM decodes the ODE-simulated signals back into 12-lead ECG space. The final loss combines the MSE term for reconstructing ECG traces with the two KL divergence penalties and an additional Jacobian-based regularization for stability. This design leverages both data-driven and mechanistic insights to produce robust, interpretable cardiac embeddings.

At each iteration, the model was shown randomly chosen mini-batches of paired ECG and phenotyping data (DEXA, iglu, fundus, ultrasound, anthropometrics). The regularization terms were not an accurate metric of model training (as they did not decrease monotonically), and so we halted training after three consecutive epochs of non-decreasing MSE test loss, averaged over the validation set. We used the Adam optimizer Kingma & Ba (2014), and an initial learning rate of 1e-4, and trained with a 90/10 train/test split. We did not utilize an additional evaluation set here, given our limited cohort size. A description of our method is shown in Figure 1, with hyperparameters given in Table 2. The baseline model was trained in the identical way, with one exception. For the baseline model, we only changed the value of $\mathrm{KL}_{\alpha_1} = 0$ setting it to zero to evaluate the model without multimodal data integration. We trained both the baseline and the integrated model on the same splits, and extracted embeddings for the held-out test set.

RESULTS

We trained cardioPRIME by sampling four seconds windows of ECG data from a diverse cohort of 3,112 participants and their deep phenotyping data.

VALIDATION

After saving and checkpointing both models, we extracted embeddings for held-out test patients from training from the ECG branch of the model, $\mathbf{z}_{\mathrm{LSTM},i}$. To validate our models, in keeping with existing work, we checked for the alignment of embeddings to known-ground truth features (output by the machine automatically) $(d = 114)$ from these same 12-lead ECGs van de Leur et al. (2022). We did this with two subsets of the ground-truth features: top associations, to assess overall alignment, and clinical features. For inference and evaluation, we set embeddings to be the

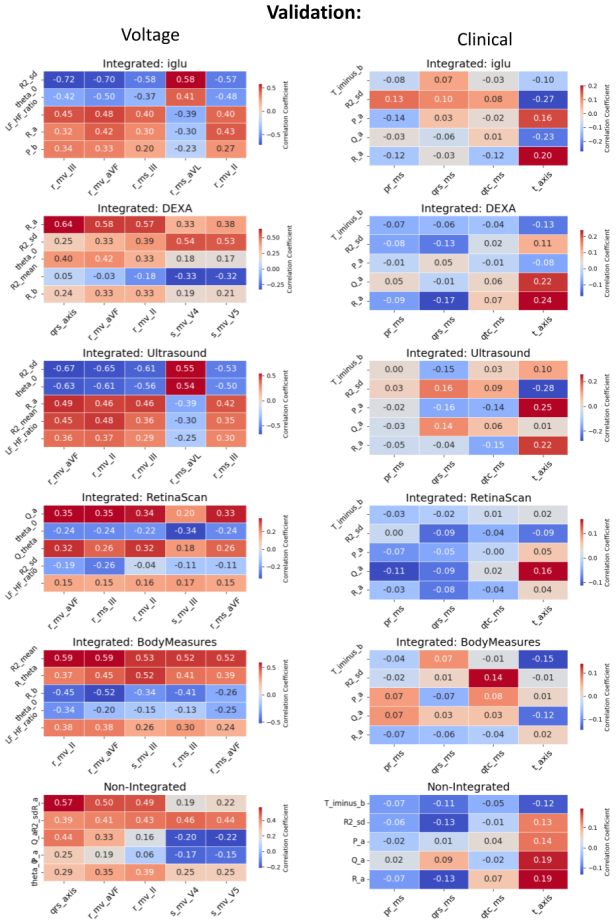

Figure 2: **Validation of ECG Latent Representations Against Ground-Truth Features.** Each heatmap row (left: the most correlated ECG features which are *Voltage* measures of the leads, right: *Clinical*-based ECG features) shows the Pearson correlation coefficients between the learned CardioPRIME embeddings and established ECG measurements. Panel rows reflect different integrated variants of CardioPRIME trained with five clinical modalities (iglu/CGM, DEXA, Ultrasound, RetinaScan, and BodyMeasures), as well as a nonintegrated baseline. Red cells indicate stronger positive correlations; blue cells indicate stronger negative correlations. The alignment between latent embeddings and these physiology-relevant ECG markers remains high across modalities, indicating that including clinical data preserves—and in some cases, strengthens—the mapping to known ECG phenomena. These findings confirm that CardioPRIME's embeddings capture key physiologic dimensions regardless of integration strategy, yet are further enriched by the additional clinical priors.

mean of their estimated distribution. Neural networks integrating this mechanistic system have been previously found to prefom similarly to black-box systems on downstream disease prediction van de Leur et al. (2022), and so we did not benchmark against a black-box deep learning model.

As demonstrated in Figure 2, the features relating to the raw ECG voltage displayed the highest correlations to learned embeddings in absolute value. Thus, we termed the former feature set (the left column of Figure 2), 'Voltage' features. To create the clinical this feature set, we selected four clinical features to use as targets for clinical validation: The PR Interval (pr_ms), which reflects atrial depolarization and AV nodal conduction, the QRS Duration (qrs_ms), which measures the time of ventricular depolarization, the corrected QT Interval (qtc_ms), which accounts for the influence of heart rate on the raw QT interval, and the T-wave Axis (t_axis), which assesses repolarization patterns. Corresponding embedding dimensions were selected based on highest average absolute correlation.

Figure 2 shows that both the nonintegrated baseline and the multimodal CardioPRIME embeddings exhibit strong correlations with established ECG features (e.g., PR interval, QRS duration). Importantly, including clinical data does not disrupt the physiological alignment; in some cases, it strengthens correlations with key conduction metrics (see T-wave axis, QRS amplitude). As well, these associations matched the scale seen in existing work van de Leur et al. (2022).

Disease Clustering Results

Having validated our physiological model both with and without clinical integration, we turned to answer our main research question. We evaluated the quality of the learned embeddings from integrating CardioPRIME with each of five deep phenotyping datasets, compared to the nonintegrated

baseline. Existing work with interpretable models in health compares K-means clusters of embeddings with ground-truth disease labels Wang & Fox (2023).

Choosing the value of K
We focused on the 10 most prevalent diagnoses in each modality's subset of patients, ranging from hypertension and hypercholesterolemia to non-cardiac conditions like musculoskeletal issues or allergic rhinitis. Diagnoses were collected at both baseline and two-year follow-up appointments for all patients. Given the known comorbidities in our disease set, such as hypercholesterolemia and hypertension Ivanovic & Tadic (2015), and the sparsity of our disease labels, we chose to use half as many disease clusters as the number of tested disease states. Including an additional 'healthy' cluster yielded a final value of $k = 6$.

Using our previously validated ECG embeddings, we applied k-means clustering to the held-out test embeddings, and stored the cluster labels for each patient. We tested whether the embeddings from the integrated ECG model outperformed the baseline one over inclusion of the five separate clincal modalities.

Evaluating embedding clusters against ground-truth disease diagnoses
The evaluation was performed using the Normalized Mutual Information (NMI), Adjusted Mutual Information (AMI), Homogeneity, and Completeness metrics, in keeping with existing work Wang & Fox (2023). These metrics assess how well the learned representations separate different disease states in an unsupervised manner. NMI and AMI measure how much information about true disease labels is retained in the clusters while correcting for random chance, ensuring that improvements are meaningful rather than accidental. Homogeneity checks whether each cluster contains primarily patients with the same disease, while Completeness ensures that all patients with a given disease are grouped together rather than being split across multiple clusters. Taken together, these metrics provide a rigorous test of whether integrating clinical data leads to more biologically meaningful and diagnostically useful representations, with higher values indicating that the learned embeddings align well with real-world disease classifications and thus improve interpretability and potential clinical applicability.

For each disease, the integrated metrics were normalized relative to baseline ones, so that an entry above one indicates that the integrated model outperforms the nonintegrated one, for the same disease and sample set. Ratios below zero as well as those exceeding 100 were omitted from the table. We exceeded ratios above 100 to avoid biasing the average performance upwards by the inclusion of rare diseases.

Green cells in Figure 3 indicate that for the same disease and ECG recordings, the integrated hybrid model outperformed the nonintegrated, but otherwise identical model, on the downstream task of disease stratfication. As seen in Figure 3, most metric ratios were above one. As well, the average of this ratio over all diseases was above one for each dataset (the bottom row in each panel). This confirms the net performance boost in simple downstream task performance from our multimodal integration.

DISCUSSION

Here, we showed that the integration of multimodal clinical features at training time led to embeddings that were more physiologically meaningful and better suited for disease separation than embeddings derived from a mechanistic model trained ECG data alone. Our goals were twofold: to validate our novel deep learning model, and then to show that clinical data integration could boost the discriminative performance of our deep learning models. To quantify our the alignment of our embeddings to known ground truths, we utilized pearson ρ, studying both top associations, and ones to explicitly clinically defined ground-truth features. To quantify the benefit of our clinical data integration, we compared K means clusters of embeddings to known ground truths, across all five datasets. We conducted all assessments of integrated embeddings relative to the identical mechanistic baseline, trained without fusion, both trained and evaluated on the identical sample set.

Notably, several conditions that exhibited improved clustering (e.g., allergic rhinitis, spinal pain) are not strictly cardiac in nature. This suggests that the integrated embeddings might capture broader metabolic or autonomic dysfunction, consistent with prior findings that ECG signals can be in-

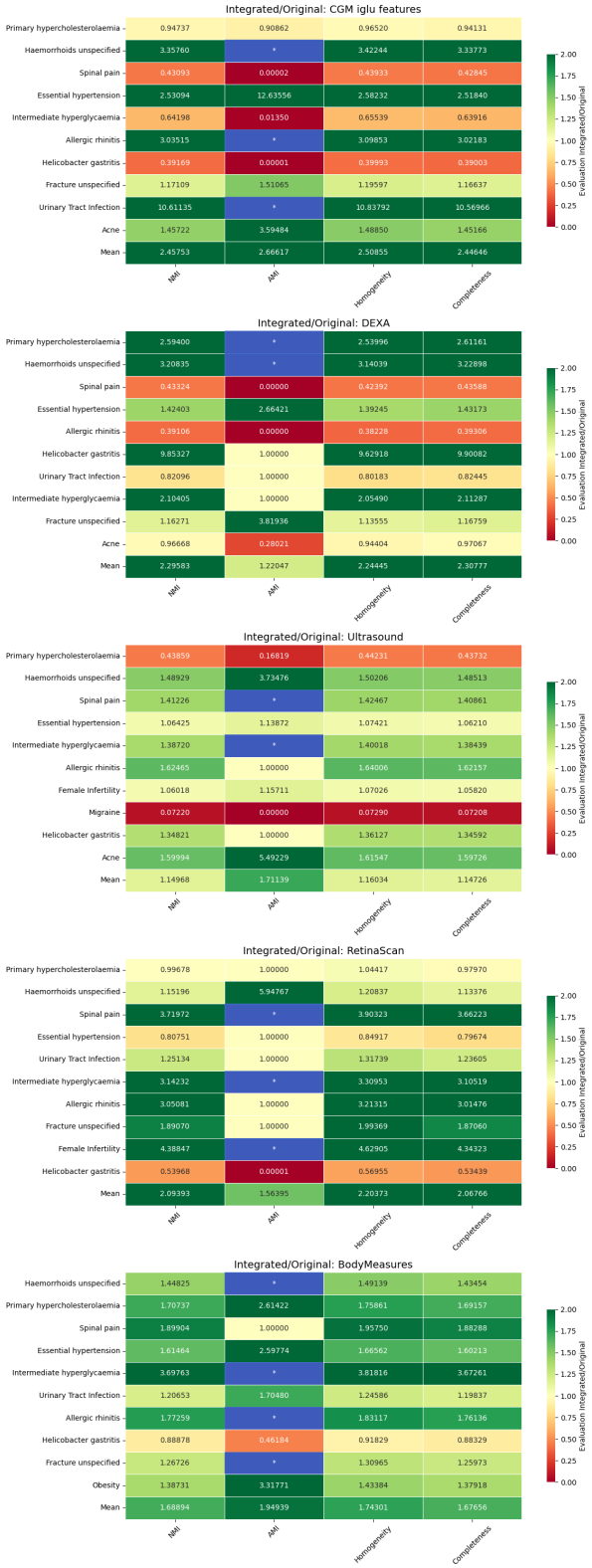

Figure 3: **Improved Disease Stratification With Multimodal Integration.** Comparison of clustering metrics (NMI, AMI, Homogeneity, and Completeness) between CardioPRIME embeddings integrated with each of five clinical phenotyping modalities (rows: CGM/iglu, DEXA, Ultrasound, RetinaScan, BodyMeasures) and the nonintegrated baseline. Each cell shows the ratio of the integrated model's performance relative to the baseline for one of the top 10 most prevalent diseases in that modality's cohort. A ratio > 1 indicates better performance under integration. Notably, even noncardiac conditions (e.g., allergic rhinitis, spinal pain, infertility) demonstrate improved clustering under CardioPRIME, highlighting the method's generality. Averaged across diseases and modalities, integration consistently enhances unsupervised disease separation. These results underscore the key message that weaving physiologic priors with multimodal clinical data yields more discriminative and clinically relevant ECG embeddings, even in the face of heterogeneous disease profiles.

fluenced by overall health status Anbalagan et al. (2023). Because our approach retains explicit ODE-based parameters (e.g., wave amplitudes, intervals), we can interpret each latent dimension in physiological terms. Clinicians or researchers can inspect whether certain parameter shifts (e.g.,

R-wave amplitude) correlate with comorbid conditions. This interpretability advantage may be particularly valuable for risk stratification in clinical workflows.

Lastly ,we selected $K = 6$ for clustering based on preliminary experiments that evaluated the effect of varying K on clustering performance. While larger values of K provided finer-grained clusters, they did not universally improve disease separation. Conversely, smaller values of K risked collapsing clinically distinct subpopulations into the same cluster. Setting $K = 6$ offered a balance between capturing disease-specific variability and maintaining cluster interpretability. Despite these positive results, there were a few diseases where the performance differences between integrated and nonintegrated embeddings were less pronounced.

CONCLUSION

We have demonstrated that integrating multimodal clinical features during training yields embeddings that are both more physiologically meaningful and more effective for disease separation than those derived solely from ECG-based mechanistic models. These results highlight the potential of hybrid mechanistic-deep learning models to enhance clinical utility in downstream tasks, such as disease classification and risk stratification.

The significance of these findings is twofold. First, they underscore the value of embedding physiological constraints into deep learning architectures, which can improve model interpretability and stability. Second, they demonstrate how baseline clinical features—traditionally disconnected from the raw ECG signal—can amplify the discriminative power of learned representations for downstream tasks, including disease stratification and risk assessment. Beyond ECG, similar integration strategies could be extended to other physiological signals or imaging modalities with well-established mechanistic foundations, towards a unified, interpretable foundation model for patient health.

MEANINGFULNESS STATEMENT

Contrary to self-supervised-learning paradigms, medical data are not corpora of meaningless bytes - tests are rooted in decades of gold-standard medical research. We propose a paradigm shift for representation models. Rather than benchmark learned features against tabular medical data, we show that including novel multimodal deep phenotyping data at training-time in an interpretable neural ordinary differential equation for electrocardiograms yields embeddings that are more useful for disease stratification than those from the same model, trained without multimodal integration. We show this over 5 multimodal datasets, on held-out test patients. Our results have useful implications for the future of interpretable mechanistic models.

ADDITIONAL TECHNICAL DETAILS

Parameter	Range	Description
P_a	[-1.7, 1.7]	Amplitude of P wave
P_theta	[-70°π/180, -50°π/180]	Relative time of P wave
P_b	[0.24, 0.35]	Scale of P wave
Q_a	[-6.5, -1.5]	Amplitude of Q wave
Q_theta	[-25°π/180, -5°π/180]	Relative time of Q wave
Q_b	[0.08, 0.12]	Scale of Q wave
R_a	[15, 55]	Amplitude of R wave
R_theta	[-20°π/180, 20°π/180]	Relative time of R wave
R_b	[0.09, 0.11]	Scale of R wave
S_a	[-0.8, -0.3]	Amplitude of S wave
S_theta	[5°π/180, 25°π/180]	Relative time of S wave
S_b	[1e-16, 0.1]	Scale of S wave
T_plus_a	[0.5, 0.9]	Amplitude of T_plus wave
T_plus_theta	[80°π/180, 120°π/180]	Relative time of T_plus wave
T_plus_b	[0.3, 0.5]	Scale of T_plus wave
T_iminus_a	[0.2, 0.9]	Amplitude of T_iminus wave
T_iminus_theta	[130°π/180, 150°π/180]	Relative time of T_iminus wave
T_iminus_b	[0.15, 0.25]	Scale of T_iminus wave
RR1_mean	[0.08, 0.12]	Mean of LF
RR1_sd	[0.008, 0.012]	Standard deviation of LF
RR2_mean	[0.23, 0.27]	Mean of HF
RR2_sd	[0.008, 0.012]	Standard deviation of HF
LF_HF_ratio	[0.4, 0.6]	Ratio of LF to HF
theta_0	[Not specified]	Initial relative time
z0_0	[Not specified]	Initial state

Table 1: Parameters, ranges, and their descriptions. LF/HF = Low/High frequency RR interval spectral component. Ranges are taken from McSharry et al. (2003) and Cli (2006)

The diversity of the constraint functions, a weighted combination of the sigmoid, hyperbolic tangent, and natural logarithm functions, was crucial the latent space to encode patient health states with sufficient resolution. While the first KL divergence term is used for clinical data integration, we also found that including a second KL divergence term, which focuses on keeping the parameters in their respective physiologically meaningful ranges, was important for learning meaningful representations. For each parameter, the mean of this distribution was set to the unconstrained mean of the range in Table 1, and we used a global $\sigma = 1$.

Our model was novel for two reasons: its small size, as well as the fusion of ECG and clinical date in the embedding domain, as opposed to a classical conditional VAE. A standard way to integrate baseline health data into a model would be to condition learned ECG representations on these separate modalities. However, in this case, feature-derived embeddings would not be separable from those based on ECG. As a result, any improvements found in the embeddings from the integrated model over the baseline one could be attributed to the richness of the included tabular data. We therefore constructed our model to match representations from ECG and clinical phenotypes that were generated separately from one another.

In the time-series domain, variational autoencoder (VAE) models have been shown to work well, especially with data from ECG Jang et al. (2021); Kuznetsov et al. (2021); Liu et al. (2020). Beyond the time domain, VAEs are also well suited for the integration of multimodal clinical data Yun et al. (2024). As we were dealing with distributions in the latent space, the natural approach was to use KL divergence to minimize the distance in learned distributions (Gaussian) from clinical data and ECGs. The result is similar to using learned priors Rezende & Mohamed (2016).

Table 2: Architectural and Hyperparameter Configuration

Component	Details/Choices	Description
Architecture Details		
Encoder	Bidirectional LSTM (hidden size: 128, layers: 1)	Encodes time-series data into a latent representation.
Baseline Projection	Two fully - connected layers with GELU activation	Projects baseline features into the latent space.
Decoder	LSTM (hidden size: 128, layers: 1) with Conv1D projection	Generates outputs from decoder hidden states.
Hyperparameters		
`hidden_size`	128	Size of hidden layers in encoder and decoder.
`num_layers`	1	Number of layers in encoder and decoder.
`dt`	$\ln(1)$	Initial value for time step for dynamics simulation.
`encoder_dropout_prob`	0	Dropout probability for the encoder.
`temperature`	1	Initial value for parameter scaling.
`mse_alpha`	100	Weight for mean squared error loss.
`jac_alpha`	5×10^{-2}	Weight for Jacobian loss.
`length_window`	4000	Length of simulation window.
`lr`	1×10^{-4}	Learning rate for optimization.
`batch_size`	20	Batch size for training.
`epochs`	250	Maximum number of training epochs (not-reached).
`kl_alpha1`	2	Weight for learning tabular data KL divergence tern.
`kl_alpha2`	1.5	Weight for physiological plausibility KL divergence.

DATA COLLECTION AND AVAILABILITY

Data in this project is part of the The Human Phenotype Project and is accessible to researchers from other academic institutions at `https://humanphenotypeproject.org/data-access`.

We set the data collection period to be from the beginning of the study, January 2019 to January 2025. For each of the features included in the The Human Phenotype Project, we kept only the latest of multiple entries; removing outliers from the data by clipping it to five standard deviations of the mean. Lastly, we excluded features with less than 2000 entries, and those that were highly unbalanced ($> 95\%$ frequency for any individual value). We only used baseline visit data for each subject, from their intake appointment. All ECGs were all sampled at a rate of 1000 hz, for 10 seconds minimum, before applying a 50Hz AC noise filter and an EMG muscle noise filter at 35 Hz. The baseline filter was set to on, and we used 16 bit resolution. For each patient, we selected one four second interval across 12 channels from their intake visit, for training. We excluded automatically-detected changes related to heart rate, artifacts (identified by focal changes in only part of the leads), and lead misplacement.

Acknowledgements: I thank members of the Segal Lab lab for useful discussions. E.S. is supported by the Crown Human Genome Center; Larson Charitable Foundation New Scientist Fund; Else Kroener Fresenius Foundation; White Rose International Foundation; Ben B. and Joyce E. Eisenberg Foundation; Nissenbaum Family; Marcos Pinheiro de Andrade and Vanessa Buchheim; Lady Michelle Michels; Aliza Moussaieff; and grants funded by the Minerva foundation with funding from the Federal German Ministry for Education and Research and by the European Research Council.

Declaration of Interests: H.R. is an employee of Pheno AI, Ltd, a biomedical data science company from Tel Aviv, Israel. E.S. is a paid consultant to Pheno AI. The rest of the authors declare no competing interests.

Ethics Statement: The Weizmann Institute of Science review board (IRB) approved the study and its protocols. All identifying details of the participants were erased prior to statistical analysis, so informed consent was waived by the IRB. All participants had full knowledge of data handling, storage, and sharing methods. This information was given to all participants, and is in agreement with the data privacy and protection policy of the Weizmann Institute

REFERENCES

Advanced methods and tools for ECG data analysis. Artech House engineering in medicine biology series. Artech House, Boston London, 2006. ISBN 978-1-58053-966-1.

Thivya Anbalagan, Malaya Kumar Nath, D. Vijayalakshmi, and Archana Anbalagan. Analysis of various techniques for ecg signal in healthcare, past, present, and future. *Biomedical Engineering Advances*, 6:100089, November 2023. ISSN 2667-0992. doi: 10.1016/j.bea.2023.100089.

Steven Broll, Jacek Urbanek, David Buchanan, Elizabeth Chun, John Muschelli, Naresh M. Punjabi, and Irina Gaynanova. Interpreting blood glucose data with r package iglu. *PLOS ONE*, 16(4): e0248560, April 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0248560.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. (arXiv:1806.07366), December 2019. doi: 10.48550/arXiv.1806.07366. URL http://arxiv.org/abs/1806.07366. arXiv:1806.07366 [cs].

Nathaniel Diamant, Erik Reinertsen, Steven Song, Aaron Aguirre, Collin Stultz, and Puneet Batra. Patient contrastive learning: a performant, expressive, and practical approach to ecg modeling. 2021. doi: 10.48550/ARXIV.2104.04569. URL https://arxiv.org/abs/2104.04569.

Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam M. Oberman. How to train your neural ode: the world of jacobian and kinetic regularization. (arXiv:2002.02798), June 2020. doi: 10.48550/arXiv.2002.02798. URL http://arxiv.org/abs/2002.02798. arXiv:2002.02798 [stat].

Sam F. Friedman, Shaan Khurshid, Rachael A. Venn, Xin Wang, Nate Diamant, Paolo Di Achille, Lu-Chen Weng, Seung Hoan Choi, Christopher Reeder, James P. Pirruccello, Pulkit Singh, Emily S. Lau, Anthony Philippakis, Christopher D. Anderson, Mahnaz Maddah, Puneet Batra, Patrick T. Ellinor, Jennifer E. Ho, and Steven A. Lubitz. Unsupervised deep learning of electrocardiograms enables scalable human disease profiling. *npj Digital Medicine*, 8(1):1–13, January 2025. ISSN 2398-6352. doi: 10.1038/s41746-024-01418-9.

Tomer Golany, Daniel Freedman, and Kira Radinsky. Simgans: Simulator-based generative adversarial networks for ecg synthesis to improve deep ecg classification. (arXiv:2006.15353), June 2020. doi: 10.48550/arXiv.2006.15353. URL http://arxiv.org/abs/2006.15353. arXiv:2006.15353 [eess].

Tomer Golany, Daniel Freedman, and Kira Radinsky. Ecg ode-gan: Learning ordinary differential equations of ecg dynamics via generative adversarial learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(11):134–141, May 2021. ISSN 2374-3468. doi: 10.1609/aaai.v35i1.16086. URL https://ojs.aaai.org/index.php/AAAI/article/view/16086.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8): 1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735.

Branislava Ivanovic and Marijana Tadic. Hypercholesterolemia and hypertension: Two sides of the same coin. *American Journal of Cardiovascular Drugs: Drugs, Devices, and Other Interventions*, 15(6):403–414, December 2015. ISSN 1179-187X. doi: 10.1007/s40256-015-0128-1.

Jong-Hwan Jang, Tae Young Kim, Hong-Seok Lim, and Dukyong Yoon. Unsupervised feature learning for electrocardiogram data using the convolutional variational autoencoder. *PLOS ONE*, 16(12):e0260612, December 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0260612.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2014. doi: 10.48550/ARXIV.1412.6980. URL https://arxiv.org/abs/1412.6980.

Dani Kiyasseh, Tingting Zhu, and David A. Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients. (arXiv:2005.13249), May 2021. doi: 10.48550/arXiv.2005.13249. URL http://arxiv.org/abs/2005.13249. arXiv:2005.13249 [cs].

V. V. Kuznetsov, V. A. Moskalenko, D. V. Gribanov, and Nikolai Yu Zolotykh. Interpretable feature generation in ecg using a variational autoencoder. *Frontiers in Genetics*, 12, April 2021. ISSN 1664-8021. doi: 10.3389/fgene.2021.638191. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.638191/full.

Ori Linial, Neta Ravid, Danny Eytan, and Uri Shalit. Generative ode modeling with known unknowns. In *Proceedings of the Conference on Health, Inference, and Learning*, pp. 79–94, April 2021. doi: 10.1145/3450439.3451866. URL http://arxiv.org/abs/2003.10775. arXiv:2003.10775 [stat].

Han Liu, Zhengbo Zhao, Xiao Chen, Rong Yu, and Qiang She. Using the vq-vae to improve the recognition of abnormalities in short-duration 12-lead electrocardiogram records. *Computer Methods and Programs in Biomedicine*, 196:105639, November 2020. ISSN 0169-2607. doi: 10.1016/j.cmpb.2020.105639.

Kaden McKeen, Laura Oliva, Sameer Masood, Augustin Toma, Barry Rubin, and Bo Wang. Ecg-fm: An open electrocardiogram foundation model. (arXiv:2408.05178), August 2024. doi: 10.48550/arXiv.2408.05178. URL http://arxiv.org/abs/2408.05178. arXiv:2408.05178 [cs].

P.E. McSharry, G.D. Clifford, L. Tarassenko, and L.A. Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3):289–294, March 2003. ISSN 1558-2531. doi: 10.1109/TBME.2003.808805. URL https://ieeexplore.ieee.org/document/1186732.

Andrew C. Miller, Nicholas J. Foti, and Emily B. Fox. Breiman's two cultures: You don't have to choose sides. (arXiv:2104.12219), April 2021. doi: 10.48550/arXiv.2104.12219. URL http://arxiv.org/abs/2104.12219. arXiv:2104.12219 [stat].

Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. (arXiv:1505.05770), June 2016. doi: 10.48550/arXiv.1505.05770. URL http://arxiv.org/abs/1505.05770. arXiv:1505.05770 [stat].

Matteo Salvador, Marina Strocchi, Francesco Regazzoni, Christoph M. Augustin, Luca Dede', Steven A. Niederer, and Alfio Quarteroni. Whole-heart electromechanical simulations using latent neural ordinary differential equations. *npj Digital Medicine*, 7(1):1–8, April 2024. ISSN 2398-6352. doi: 10.1038/s41746-024-01084-x.

Rutger R van de Leur, Max N Bos, Karim Taha, Arjan Sammani, Ming Wai Yeung, Stefan van Duijvenboden, Pier D Lambiase, Rutger J Hassink, Pim van der Harst, Pieter A Doevendans, Deepak K Gupta, and René van Es. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. *European Heart Journal - Digital Health*, 3(3):390–404, September 2022. ISSN 2634-3916. doi: 10.1093/ehjdh/ztac038. URL https://doi.org/10.1093/ehjdh/ztac038.

Ke Alexander Wang and Emily B. Fox. Interpretable mechanistic representations for meal-level glycemic control in the wild. In Stefan Hegselmann, Antonio Parziale, Divya Shanmugam, Shengpu Tang, Mercy Nyamewaa Asiedu, Serina Chang, Tom Hartvigsen, and Harvineet Singh (eds.), *Proceedings of the 3rd Machine Learning for Health Symposium*, volume 225 of *Proceedings of Machine Learning Research*, pp. 607–622. PMLR, 10 Dec 2023. URL https://proceedings.mlr.press/v225/wang23a.html.

Yakir Yehuda and Kira Radinsky. Ordinary differential equations for enhanced 12-lead ecg generation. (arXiv:2409.17833), September 2024. doi: 10.48550/arXiv.2409.17833. URL http://arxiv.org/abs/2409.17833. arXiv:2409.17833 [cs].

Taedong Yun, Justin Cosentino, Babak Behsaz, Zachary R. McCaw, Davin Hill, Robert Luben, Dongbing Lai, John Bates, Howard Yang, Tae-Hwi Schwantes-An, Yuchen Zhou, Anthony P. Khawaja, Andrew Carroll, Brian D. Hobbs, Michael H. Cho, Cory Y. McLean, and Farhad Hormozdiari. Unsupervised representation learning on high-dimensional clinical data improves genomic discovery and prediction. *Nature Genetics*, 56(8):1604–1613, August 2024. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-024-01831-6. URL https://www.nature.com/articles/s41588-024-01831-6.

Yukun Zhou, Siegfried K. Wagner, Mark A. Chia, An Zhao, Peter Woodward-Court, Moucheng Xu, Robbert Struyven, Daniel C. Alexander, and Pearse A. Keane. Automorph: Automated retinal vascular morphology quantification via a deep learning pipeline. *Translational Vision Science Technology*, 11(7):12, July 2022. ISSN 2164-2591. doi: 10.1167/tvst.11.7.12.