The University of Kansas Medical Center

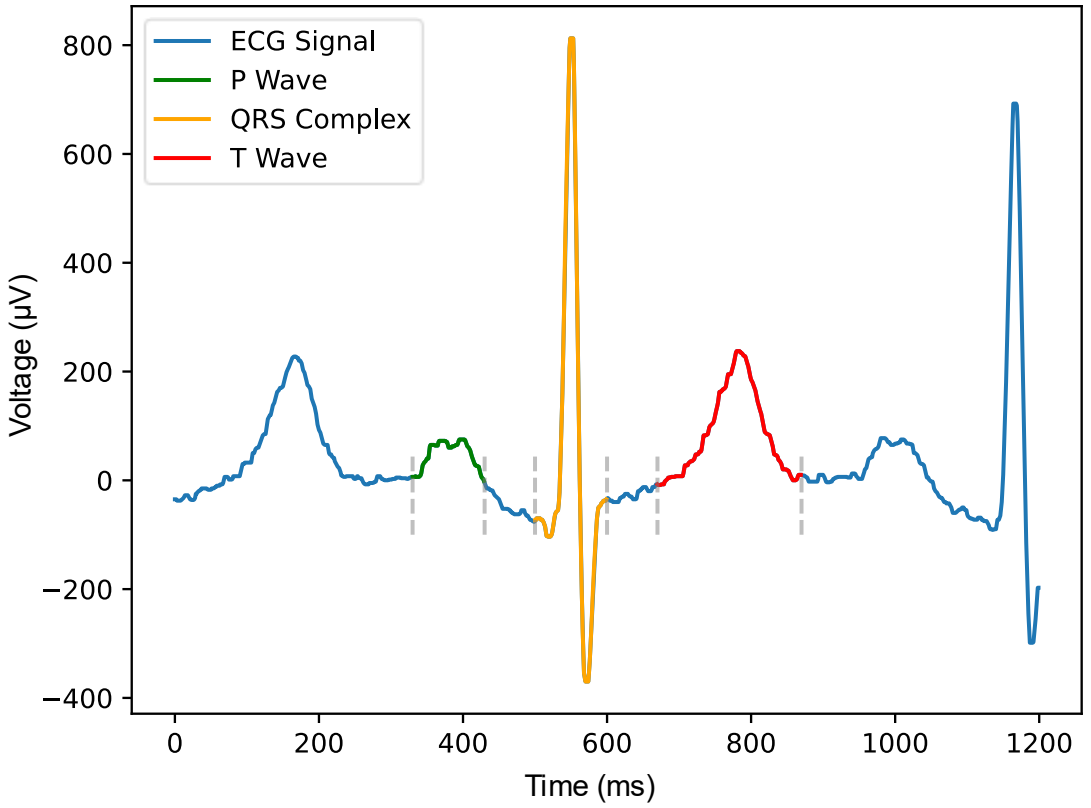
ECG Latent Feature Extraction with Autoencoders for Downstream Prediction Tasks

Chris Harvey
Sumaiya Shomaji
Zijun Yao
Amit Noheria

ECG Data

The electrocardiogram is a 10-second, 12-lead recording sampled at 500 Hz, producing about **60,000 data points** and capturing 8–17 heartbeats. Each cycle includes the P wave (atrial depolarization, green/left), QRS complex (ventricular depolarization, orange/center), and T wave (ventricular repolarization, red/right), which differ dramatically in duration, amplitude, and frequency.

ECG signals are highly skewed due to intervals of electrical inactivity, and **morphology varies widely across individuals**—ranging from smooth to notched or triphasic complexes—reflecting differences in anatomy, physiology, lead placement, and disease.



Background on the problem

Deep learning performs well on ECG only when supported by large datasets, because a 10-second, 12-lead ECG contains 60,000 high-dimensional datapoints with substantial morphological variability and skewed distributions.

As data complexity increases, the amount of data needed to model it also increases. The complexity of ECG signals poses challenges for Deep Learning (DL) due to high variance in wave morphology, skewed data distribution, and temporal volatility. Hence, we **need** a solution to the ECG data complexity problem in smaller datasets.

Smaller datasets—such as those involving rare conditions or invasive procedures e.g., post-ablation atrial fibrillation (A-Fib)—**are insufficient for DL to generalize without overfitting**.

Variational autoencoders address this problem by learning **compact representations** and reconstructing signals, **enabling robust feature extraction** when data volume is limited. Our goal is to exploit such representations to build reliable predictive models despite restricted sample sizes.

The UNIVERSITY of KANSAS Medical CENTER

Our Solution. The S-VAE Model

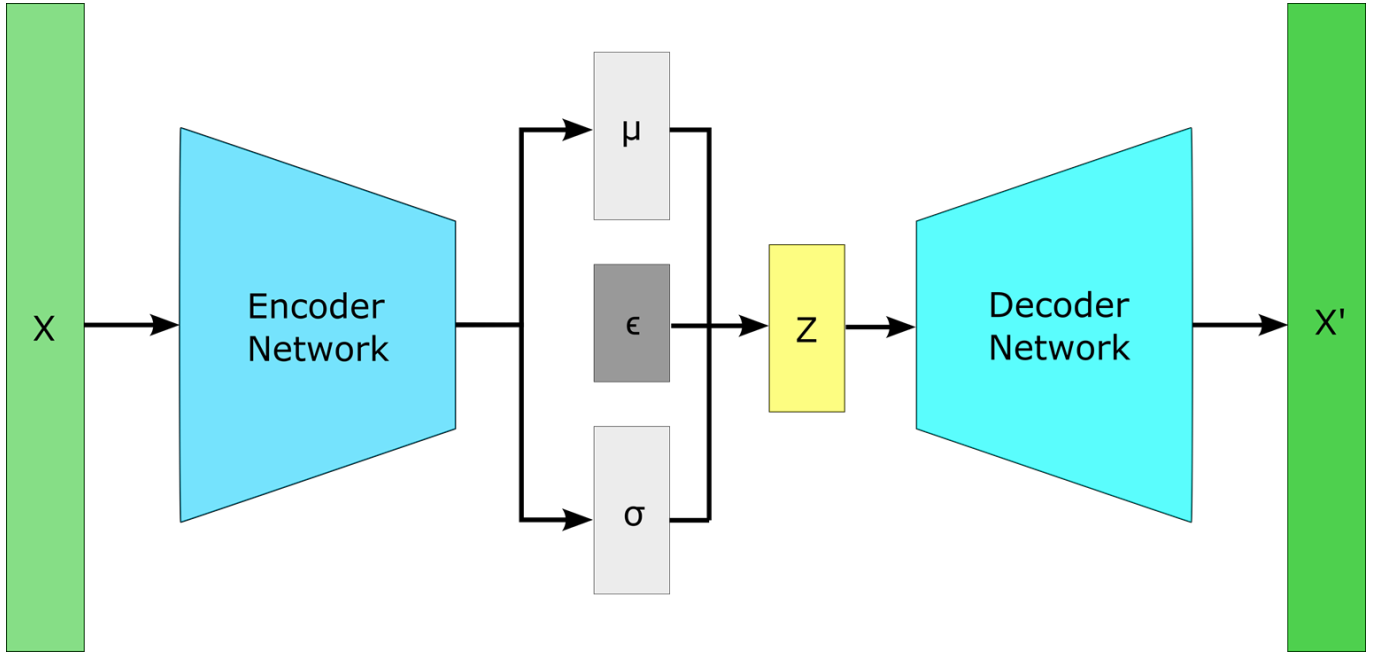
We propose a VAE-based framework with three novel variants—Stochastic Autoencoder (SAE), Cyclical βVAE (Cβ-VAE), and Annealed βVAE (Aβ-VAE)—to optimize ECG latent representations for high-fidelity signal reconstruction and improved predictive performance, especially with **limited-size training datasets**.

These variants balance reconstruction fidelity and latent space regularization, enabling DL on limited ECG datasets and facilitate integration with simpler algorithms like tree-based models, offering an alternative to standard DL methods like Convolutional Neural Networks (CNNs).

We trained these novel VAEs on >1 million ECGs to represent representative beat x-y-z-lead ECG into 30 encodings. A compression of 60,000 data points to 30. VAE extracted encodings were used to train LGBM models to classify labels from conventional ECGs.

Model Pipeline

All VAE models featured an encoder-decoder structure built with Convolutional Neural Networks (CNNs).

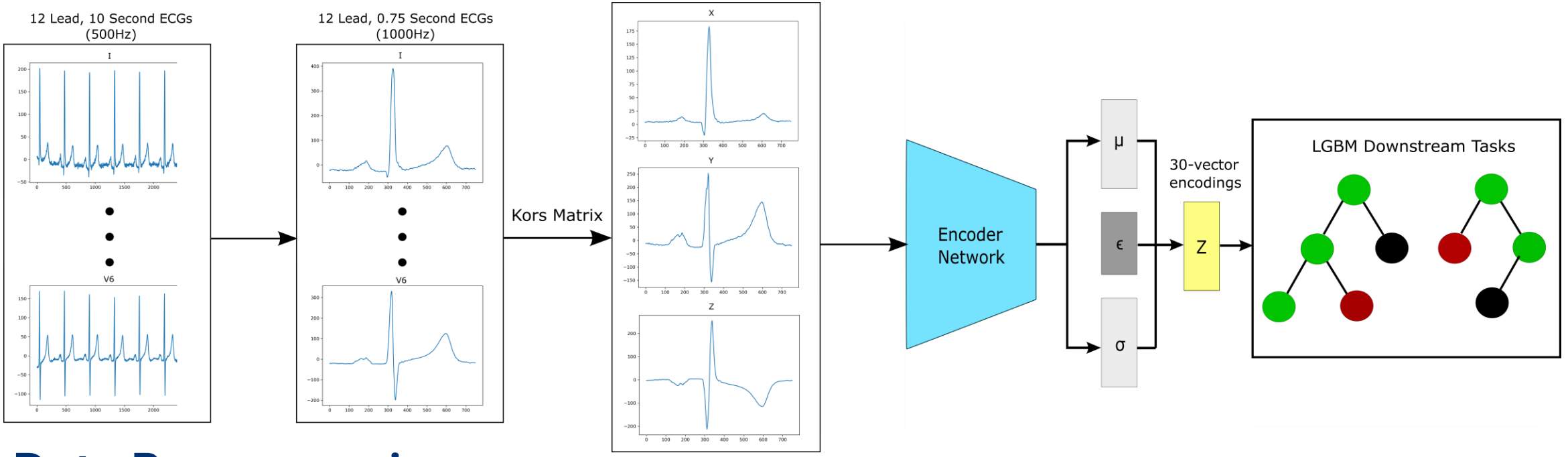


The encoder included four 2D convolutional layers (filters: 256, 256, 512, 512) with a filter width of 9, stride of 2, TanH activations, and batch normalization, followed by two fully connected layers with L2 regularization (0.01) and dropout (0.25).

The latent variable z was sampled from a Gaussian distribution (except the SAE, which is a stochastic distribution) via the reparameterization trick.

The decoder mirrored the encoder with two fully connected layers and four transpose convolutional layers (filters: 512, 256, 128, 3) to reconstruct the signal. Models were trained on ~1.1 million ECGs using TensorFlow, Adam optimizer (learning rate: 0.000001), 50 epochs, and a batch size of 32, on a GeForce RTX 3090 with 128 GB RAM and an AMD Ryzen 9 3900XT CPU, though inference is lightweight and hardware agnostic.

The UNIVERSITY of KANSAS Medical CENTER



Data Preprocessing

We reduced 10-sec ECG recordings to a 750ms representative beat, created by the Philips IntelliSpaceECG system to produce a mean average beat from the 10 second signal. This was notch filtered (60 Hz), Butterworth band-pass filtered (0.05Hz low, 60Hz high), and median filtered (rolling window width of 11). The use of the representative beat captures key morphological features while shrinking data size by eliminating the redundant multiple cardiac cycles.

Then, using Kors's conversion matrix, we transformed the eight independent ECG leads (I, II, V1-V6) into three orthogonal X (right to left), Y (cranial to caudal), and Z (anterior to posterior) leads, reducing the 1,000Hz 120,000-datapoint 10-sec 12-lead ECG to a 2250-datapoint 750ms 3-lead ECG as shown in Figure 1.

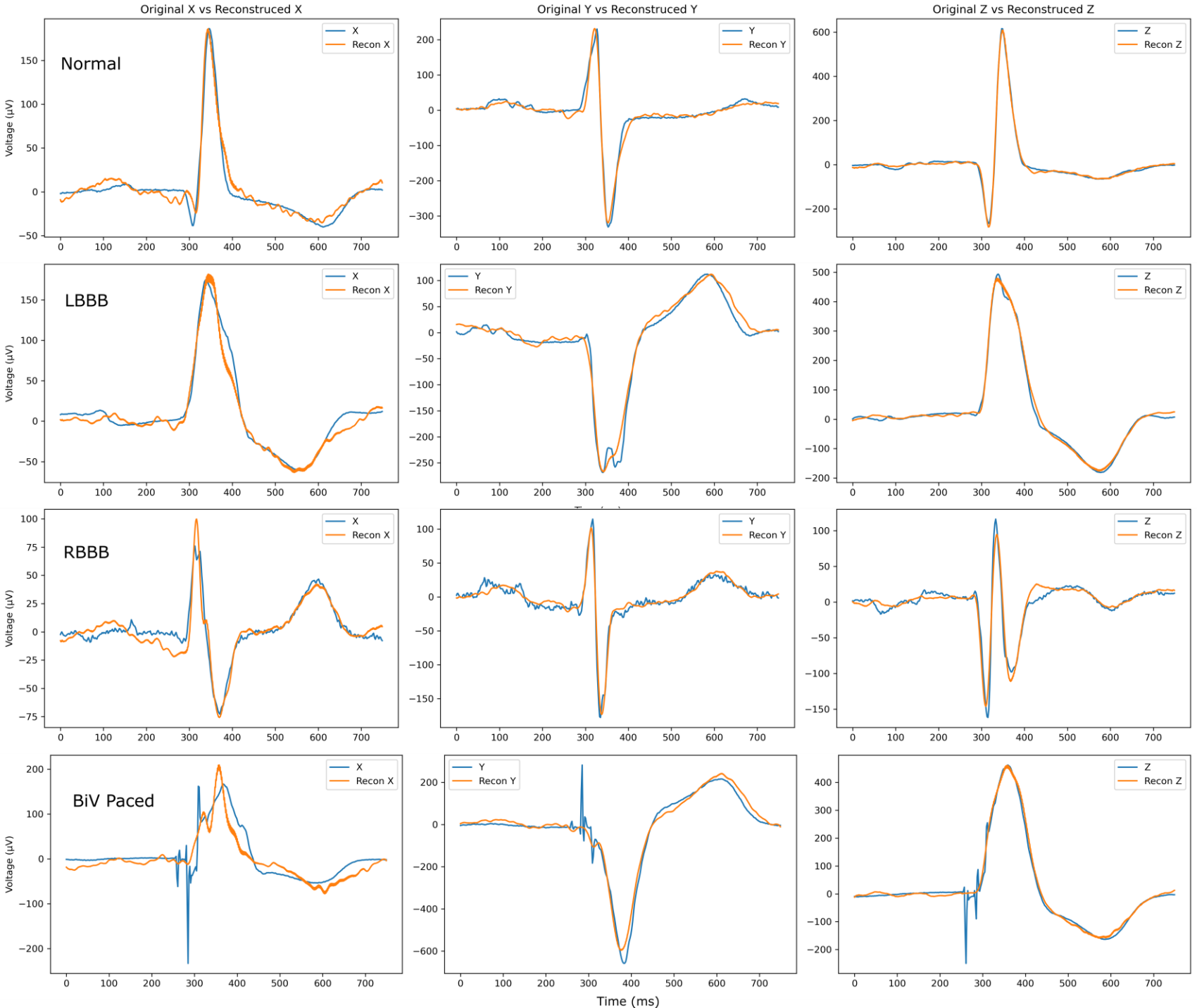
KL Loss Beta Values Used

In the Cβ-VAE, we adopt a cyclical annealing schedule in which β varies from 0 to 5 across training, effectively toggling the model between three regimes: pure autoencoder (β=0), standard VAE (β=1), and β-VAE (β>1). At β=0, the loss consists solely of reconstruction error; at β=1, the model incorporates a conventional KL penalty; and at β>1, the KL term is amplified to prioritize disentanglement. Whereas the original cyclical annealing work cycled β between 0 and 1 with a hard reset, our approach spans 0–5 without reset, rising over 10 epochs and decaying over the next 10, yielding a 20-epoch cycle. This configuration accelerates convergence, enabling usable reconstructions within approximately 10 epochs.

The Aβ-VAE applies a reverse annealing strategy in which β begins at 10 and declines to 0 over 50 epochs. **This allows the model to initially emphasize latent space structure and disentanglement before shifting toward reconstruction fidelity**, resulting in superior generative output.

By contrast, the SAE fixes β=0, **removing the KL term entirely**. This produces a stochastic encoder trained only via reconstruction loss, learning whatever latent distribution best minimizes error rather than approximating a Gaussian prior. While the classical VAE framework targets Gaussian latent structure for generation, the SAE strategically abandons that constraint to maximize reconstruction quality—**especially beneficial for downstream ECG tasks**—while retaining some stochastic advantages of VAEs.

The UNIVERSITY of KANSAS Medical CENTER



Reconstruction Quality

Reconstructed ECGs from the VAE encodings.
Blue=Original
Orange=Reconstructed

Model	1st 250 ms (p wave) MAE (μV) Avg ± SD	2nd 250 ms (QRS) MAE (μV) Avg ± SD	3rd 250 ms (T wave) MAE (μV) Avg ± SD	Full signal MAE MAE (μV) Avg ± SD	Full signal MSE (μV^2) Avg ± SD	Full signal DTW Avg ± SD
PCA	19.1±6.5	29.5±7.3	22.7±7.7	24.0±5.0	1842.9±840.0	667.9±218.7
AE	**11.2±3.0**	23.2±5.5	**12.8±3.7**	15.8±3.1	739.8±316.1	313.7±94.0
SAE	11.4±2.7	31.7±8.1	12.7±3.4	18.7±3.7	1131.3±514.8	387.3±131.8
VAE	11.9±2.7	28.9±6.9	12.8±3.3	17.9±3.5	996.2±433.2	361.2±115.5
β-VAE	11.5±3.0	23.6±5.6	13.1±3.7	16.2±3.2	755.6±325.6	317.5±94.8
Aβ-VAE	**11.2±3.0**	**22.6±5.3**	**12.8±3.8**	**15.7±3.2**	**701.6±304.8**	**308.1±92.1**
cβ-VAE	12.0±2.6	31.9±6.6	14.0±3.8	19.3±3.4	1202.7±491.3	400.6±130.3

Model	X Signal MAE (μV)	Y Signal MAE (μV)	Z Signal MAE (μV)	X Signal DTW	Y Signal DTW	Z Signal DTW
PCA	25.8±21.4	25.6±20.7	23.5±21.0	751.9±856.4	715.8±778.0	647.2±766.0
AE	16.3±13.4	16.9±13.1	16.5±13.2	338.0±412.4	337.0±384.9	332.6±390.8
SAE	19.7±12.8	21.1±12.7	17.1±11.9	429.5±404.8	455.9±406.5	340.0±362.3
VAE	18.9±14.0	20.5±14.4	16.4±11.8	400.4±413.1	419.5±400.1	327.9±349.7
β-VAE	16.9±14.0	17.8±14.0	**15.9±12.6**	346.5±413.8	350.4±390.2	321.4±376.6
Aβ-VAE	**16.1±13.5**	**16.9±12.9**	16.0±13.1	**334.3±411.2**	**335.8±384.4**	**321.7±381.1**
cβ-VAE	20.6±14.9	22.0±15.2	17.8±11.5	438.2±425.6	462.3±422.5	365.5±358.2

MAE, MSE, and DTW for different models for representative beat X, Y, Z-lead ECG reconstructions (N=1,065,368)

Experiments and Results

We evaluated seven dimensionality-reduction approaches for ECG data: PCA, AE, SAE, VAE, β-VAE, Cβ-VAE, and Aβ-VAE. The SAE, Cβ-VAE, and Aβ-VAE represent novel VAE-based designs incorporating mechanisms intended to address the unique demands of ECG representation learning.

Incremental PCA from scikit-learn was configured to produce 30 components, matching the 30-dimensional latent encodings of the VAE models to ensure methodological parity. All autoencoder variants were trained using a shared network architecture, differentiated only by loss-function specification as outlined in later sections.

For downstream evaluation, we selected prediction of reduced left ventricular ejection fraction (LVEF ≤ 35%), the principal defining feature of heart failure with reduced ejection fraction. Our benchmark was a CNN trained directly on 10-second, 12-lead ECG signals, based on Mayo Clinic's CNN architecture.

We then compared this baseline to Light Gradient Boosted Machine (LGBM) models trained on latent encodings from each autoencoder. For LGBM, eight hyperparameters were modified from defaults: Maximum Depth (15), Colsample Bytree (0.9), Extra Trees mode (True), Top-K (100), Learning Rate (0.1), number of estimators (1,000,000 with early stopping), and regularization parameters α and λ (0.95).

Model	QRS Duration (ms)		$\text{Amplitude}_{\text{QRS-3D}}$ (μV)		$\text{VTI}_{\text{QRS-3D}}$ (μVs)	
	MAE ± SD	R^2	MAE ± SD	R^2	MAE ± SD	R^2
PCA	**8.1±12.9**	**0.727**	**108.3±173.3**	**0.838**	**3.16±5.32**	**0.923**
AE	8.7±13.8	0.687	108.8±175.3	0.835	3.68±5.96	0.903
SAE	8.3±13.2	0.712	109.1±176.6	0.835	3.42±5.64	0.914
VAE	8.4±13.4	0.705	109.7±177.6	0.833	3.46±5.72	0.911
β-VAE	8.5±13.5	0.697	109.6±177.4	0.833	3.54±5.83	0.908
Aβ-VAE	8.3±13.3	0.709	107.7±175.3	0.837	3.44±5.68	0.912
cβ-VAE	8.3±13.3	0.708	110.2±177.9	0.832	3.44±5.69	0.912

Prediction in test set of ECG measurements with LGBM using representative beat X, Y, Z-lead ECG encoded variables (n=97,464)

Model	RBBB (8.06% Prevalence)		LBBB (3.99% Prevalence)	
	AUROC	Sensitivity	AUROC	Sensitivity
PCA	0.9435	0.894	0.9637	0.939
AE	0.9390	0.881	0.9618	0.938
SAE	0.9504	0.906	**0.9701**	**0.948**
VAE	0.9507	0.904	0.9688	0.950
β-VAE	0.9473	0.895	0.9689	0.949
Aβ-VAE	0.9499	0.903	0.9686	0.947
cβ-VAE	**0.9516**	**0.908**	0.9697	0.949

Prediction in test set of RBBB and LBBB with LGBM using representative beat X, Y, Z-lead ECG encoded variables (n=97,464), specificity set at 0.9.

Model	Reduced LVEF (≤35%)		LVEF, %	
	AUROC	Sensitivity	MAE ± SD	R^2
PCA	0.799	0.616	**8.86±12.15**	**0.247**
AE	0.810	0.656	9.05±12.42	0.213
SAE	**0.820**	**0.665**	8.96±12.28	0.231
VAE	0.819	0.676	8.95±12.26	0.233
β-VAE	0.812	0.663	9.05±12.39	0.217
Aβ-VAE	0.818	0.666	8.96±12.29	0.229
cβ-VAE	0.820	0.675	8.97±12.28	0.231
ECG statistics	0.761	0.554	9.55±12.92	0.148

Prediction in test set of reduced LVEF with LGBM using representative beat X, Y, Z-lead ECG encoded variables in test set (n=30,554). Reduced LVEF 14.09% prevalence, specificity set at 0.9.

Performance of different machine learning models in predicting reduced LVEF (LVEF≤35%) from ECG data (Holdout Test Set: n=15,987), specificity set at 0.9.

Model	Training Sample Size	AUROC	Sensitivity
CNN_a	100% (n=143,644)	**0.909**	**0.742**
	9.5% (n=13,568)	0.630	0.177
$ResNet_b$	100% (n=143,644)	**0.892**	**0.672**
	10% (n=14,364)	0.855	0.586
	1% (n=1,436)	0.811	0.462
	0.1% (n=143)	0.705	0.281
$LGBM_b$	100% (n=143,644)	**0.901**	**0.702**
	10% (n=14,364)	0.870	0.610
	1% (n=1,436)	0.846	0.525
	0.1% (n=143)	0.761	0.361

[a] trained with raw signal data

[b] trained with summary statistics and SAE encodings.

SAEs outperform the other variants

The strong downstream performance of the SAE indicates that emphasizing reconstruction fidelity rather than latent space regularization **can be advantageous for ECG representation learning**. Unlike conventional VAE variants, the SAE optimizes exclusively for reconstruction error, enabling it to **capture ECG variability more effectively**. Notably, the inferior results of the deterministic AE underscore the contribution of stochastic sampling in enhancing generalizability.

SAE encodings **consistently outperformed** alternatives across predictive tasks while **maintaining high-quality signal reconstruction**, challenging the prevailing assumption that explicit latent regularization is universally beneficial. These findings suggest that reconstruction-centric objectives can yield encodings that are equally—and in some cases more—informative. Moreover, the broader VAE family retains the capacity to generate synthetic signal data, **providing additional utility beyond compression and prediction**.

The UNIVERSITY of KANSAS Medical CENTER

Conclusions

These findings demonstrate that VAE-derived encodings enable conventional machine learning models to achieve performance comparable to deep learning systems trained on full-resolution ECG waveforms, despite requiring far less data and computational overhead. This capability is particularly advantageous for diagnostic development in minority cohorts, rare disease contexts, or invasive clinical settings where large datasets are impractical or unavailable.

More broadly, our results show that the intrinsic complexity and heterogeneity of ECG signals can be effectively compressed via PCA and VAE frameworks for diverse downstream predictive tasks. Specifically, we show that a 10-second, 12-lead ECG comprising roughly 60,000 samples at 500 Hz can be distilled to a 30-dimension latent representation with minimal loss of clinically relevant information.

Thank you for your attention

Any Questions?

The UNIVERSITY of KANSAS Medical CENTER