Learning Sex-specific Latent Factors from Single-Lead ECG Signals Using a Variational Autoencoder

Ángela Hernández[1], Josseline Madrid[1], Carlos Sánchez[1,2], Patricia B Munroe[3], Alejandro Sanz[4], Julia Ramírez[1,2,3], Ana Mincholé[1,2]

[1]Aragón Institute of Engineering Research, ISS-Aragón, University of Zaragoza, Spain
[2]Centro de Investigación Biomédica en Red, Biomateriales, Bioingeniería y Nanomedicina, Spain
[3]William Harvey Research Institute, Queen Mary University of London, London, United Kingdom
[4]Endocrinology service, Miguel Servet University Hospital, Zaragoza, Spain

Abstract

Biological sex significantly modulates electrocardiogram (ECG) morphology, potentially confounding the identification of subtle disease-related signatures and cardiovascular risk. However, extracting sex-specific biomarkers without relying on supervised labels remains a challenge. Here, we propose an unsupervised generative framework using a Factor-Variational Autoencoder (FactorVAE) to isolate sex-specific cardiac features from Lead I-ECGs in the UK-Biobank cohort ($N = 54,773$). To prevent clinical confounding, the decoder was informed by age, body mass index, and blood pressure, reducing the need for the latent representation to encode these covariates, while biological sex was entirely withheld from model training. After optimization (MSE $= 0.037mV^2$, KL loss $= 25$ nats), 11 out of 32 latent dimensions (LDs) remained active. Statistical probing identified a single factor (z_{14}) correlated with biological sex ($|r_{\mathrm{pb}}| = 0.29$). A linear logistic regression on z_{14} achieved a 63.6% sex classification accuracy (AUC = 0.671), indicating that linearly accessible sex-related information was encoded within this single factor. Furthermore, latent traversals and Pearson correlations revealed alignment with ventricular repolarization biomarkers (T-wave and ST-segment amplitudes). These findings establish a framework to disentangle sex-specific ECG variations without explicit labels, paving the way for unbiased cardiac biomarker discovery.

1. Introduction

Sex plays a critical role in the electrophysiological substrate of the heart and arrhythmogenic risk. Documented sex-specific electrocardiographic (ECG) differences, such as longer QT intervals or narrower QRS–T angles in women [1], are essential for accurate clinical interpretation and cardiovascular (CV) risk stratification when using the ECG [2]. However, conventional hand-crafted ECG biomarkers may not capture the full non-linear complexity of the signal, where sex-induced morphological variations can mask purely arrhythmogenic patterns and hinder the detection of underlying cardiac pathologies [3].

With the advent of deep learning, Variational Autoencoders (VAEs) offer powerful unsupervised representation learning for ECG analysis [4]. A standard VAE maps high-dimensional ECG signals into a continuous lower-dimensional latent space, capturing critical features required for signal reconstruction [5]. Despite their efficacy, conventional VAEs frequently suffer from feature entanglement, where a single latent dimension (LD) might simultaneously encode multiple independent biological factors [6]. This limits interpretability and prevents the clean isolation of sex-related ECG characteristics and other underlying arrhythmogenic patterns.

To overcome this, disentangled representation learning isolates distinct explanatory factors into independent LDs. Although the β-VAE [6] enforces disentanglement via a weighted Kullback-Leibler (KL) penalty, this global constraint may compromise reconstruction fidelity as disentanglement is increased without guaranteeing statistical independence among LDs. Conversely, the Factor–VAE [7] introduces a dedicated total correlation (TC) penalty in the loss function that directly penalizes statistical dependence among LDs, reducing the trade-off between disentanglement and reconstruction accuracy [8].

In this work, we developed a FactorVAE framework designed to investigate whether sex-related ECG variability emerges as an independent latent factor. By explicitly conditioning the decoder on continuous anthropometric and clinical covariates while withholding biological sex during training, we evaluate whether sex-related variability is a dominant, learnable factor in ECG signals that can be isolated in an unsupervised manner, providing interpretable insights into sex-specific ECG characteristics.

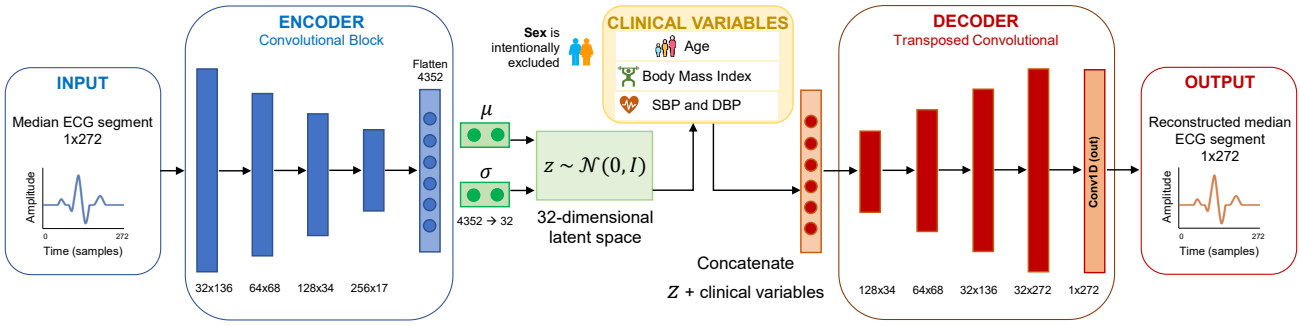

Figure 1. Proposed FactorVAE model architecture. Median single-lead ECG segments are mapped to a 32-dimensional latent space z. Clinical variables are concatenated exclusively at the decoder stage, whereas sex is intentionally excluded.

2. Materials and methods

2.1. Study population

This study used Lead I ECG recordings from the UK Biobank Exercise cohort [9] ($N = 54,773$) and matched sex, age, body mass index (BMI), and both systolic and diastolic blood pressure (SBP and DBP, respectively). ECGs were acquired during a bicycle exercise stress test, from which only the initial 15-second resting segment was analyzed, prioritizing the earliest available recording. Participants had no prior history of CV disease. The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee, and this research was conducted under application number 8256.

2.2. ECG-Preprocessing

Raw ECG signals underwent 0.3-Hz high-pass filtering, cubic spline baseline correction, and 40-Hz low-pass filtering [10]. A median heartbeat was derived per participant, R-peak aligned and cropped to 272 samples (500 Hz). Standard ECG metrics (QRS amplitude/duration, QTc, ST, and T-wave amplitudes) were extracted from these median beats for downstream interpretability analyses.

2.3. Disentangled Variational Architecture

A FactorVAE maps the Lead I ECG signals into a lower-dimensional representation 1. The encoder comprises four 1D convolutional blocks (kernel size 4, stride 2, padding 1) with 32, 64, 128, and 256 channels, producing a 256×17 feature map. This map is flattened and projected into a 32-dimensional latent space via two linear layers estimating the mean and log-variance of a Gaussian posterior, sampled through the reparameterization trick.

The decoder conditions the reconstruction on non-sex clinical characteristics by concatenating the latent vector with four covariates (age, BMI, SBP, and DBP) into a 36-dimensional input. This vector is projected back to the encoder's final shape and progressively upsampled to the original 272-sample waveform using four transposed convolutional stages and a final 1D convolution. By conditioning exclusively at the decoder stage, the encoder receives only the raw ECG waveform as input. Consequently, because the decoder is directly provided with the clinical covariates, the reconstruction loss penalizes z for encoding redundant covariate information, implicitly forcing z to filter out these confounding clinical variations. Sex is intentionally excluded from the conditioning variables so that sex-related information is not explicitly provided to the model, allowing us to evaluate whether it emerges spontaneously from the ECG signal as an independent latent factor.

The framework is trained by minimizing the reconstruction mean squared error ($\mathcal{L}_{\text{MSE}}$) and two regularization terms. The latent distribution is regularized using a capacity-controlled KL divergence objective with $C = 25$ nats and $\beta = 1$, following established methodology [11]. Independence is further enforced by a TC penalty estimated through an adversarial discriminator (two hidden layers, 16 units each) using a cross-entropy loss to distinguish between real latent vectors $z \sim q(z)$ and permuted vectors $\bar{z} \sim \bar{q}(z)$ [7]. The TC term ($\gamma = 3$) is introduced dynamically via a linear warm-up coefficient over the first 10 epochs. Detailed model architecture is illustrated in Figure 1. The overall objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \beta \left|\text{KL} - \text{C}\right| + \gamma \, \mathcal{L}_{\text{TC}}$$

The final model was trained using a standard 80/10/10 split, retaining the test set to report performance metrics.

2.4. Latent Space Analysis and Interpretability

After optimization, active LDs were identified by evaluating the individual KL divergence KL_i and filtering for

$\mathrm{KL}_i > 0.1$, as performed in [11]. To interpret the latent structure capturing sex-specific features, the active LD showing the highest point-biserial correlation (r_{pb}) with biological sex was selected. Additionally, Pearson correlation coefficients (r) were then computed between this specific LD and the subjects' clinical ECG parameters to evaluate its alignment with established sex-specific biomarkers. Next, latent traversals, modulating the isolated dimension across $[-3, +3]$ standard deviations while holding all other LDs fixed, were generated over 10 discrete steps. This visually confirms that the unsupervised model naturally captures and translates known physiological differences into specific ECG morphological variations.

Finally, to quantify the sex-specific information encoded within the selected single LD, a linear logistic regression probe was implemented. This probe evaluates the linear separability of biological sex using exclusively the isolated dimension as the input feature.

3. Results

The trained FactorVAE effectively reconstructed the main ECG morphology across the test set ($N = 8,217$), yielding a MSE of 0.037 mV^2. As illustrated in Figure 2, while the network effectively captures the dominant macro-morphology of the ECG complex (such as the R-wave peak and general T-wave orientation), it exhibits limitations in tracking low-amplitude segments and high-frequency components. This smoothing effect is a well-documented characteristic of disentangled VAEs, where heavily penalizing latent capacity trades off the capturing of residual noise in favor of statistically independent, interpretable latent factors [11].

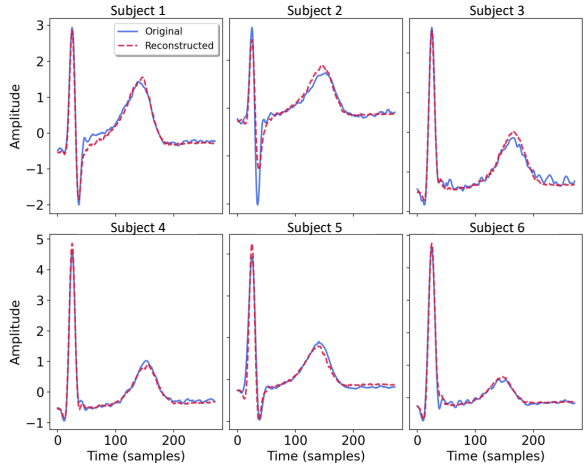

Figure 2. FactorVAE reconstruction performance. Original (blue, solid line) versus reconstructed (red, dashed line) Lead I-ECG signals across six random test subjects.

Regarding latent space optimization, the model converged to a total KL divergence of 25 nats while effectively enforcing disentanglement. Out of the 32 initial LDs, 11 dimensions were identified as active ($\mathrm{KL}_i > 0.1$). The remaining 21 dimensions were successfully naturally collapsed ($\mathrm{KL}_i \approx 0$), indicating that the TC penalty ($\gamma = 3$) forced the model to compress the cardiac information into a highly compact, non-redundant representation.

3.1. Identification of Sex-Specific Latent Structure

Among the active LDs, dimension z_{14} emerged as the primary encoder of sex-specific traits, achieving the highest point-biserial correlation with biological sex ($|r_{\mathrm{pb}}| = 0.29$). As illustrated in Figure 3 (A), statistical analysis against standard ECG biomarkers revealed that z_{14} strongly correlates with ventricular repolarization features, specifically T-wave amplitude ($|r| = 0.78$) and ST-segment amplitude ($|r| = 0.60$), demonstrating a targeted alignment with known physiological sex variations [1].

This behavior was further validated through generative latent traversals. As shown in Figure 3 (B), tracking the traversal from negative to positive latent values reveals a pronounced, gradual depression of the ST segment alongside a progressive lowering of the T-wave amplitude. Crucially, this significant alteration of the repolarization phase occurs while the underlying QRS morphology remains preserved, suggesting that z_{14} preferentially modulates repolarization-related ECG features.

Finally, the linear logistic regression probe trained on the isolated z_{14} dimension demonstrated sex-related discriminative information on the independent test set. The model achieved an accuracy of 63.6% and an Area Under the ROC Curve (AUC) of 0.671, with balanced macro-averaged precision and recall of 63.0%. Given that biological sex was completely withheld from the encoder during training, these classification metrics confirm that the unsupervised FactorVAE encoded linearly accessible sex-related ECG variation within a single latent factor.

4. Discussion and conclusions

In this work, we demonstrated that an unsupervised FactorVAE identify a latent factor associated with sex-related ECG features without using biological sex during model training, indicating that sex-related electrophysiological variability can emerge naturally within the learned latent representation. Interestingly, the isolated dimension z_{14} captured a statistically significant but moderate correlation with biological sex. We hypothesize that this moderate linear separability arises because sex-related cardiovascular traits remain partially distributed across other active LDs. This aligns with a fundamental limitation of unsupervised disentanglement approaches: while they successfully iso-

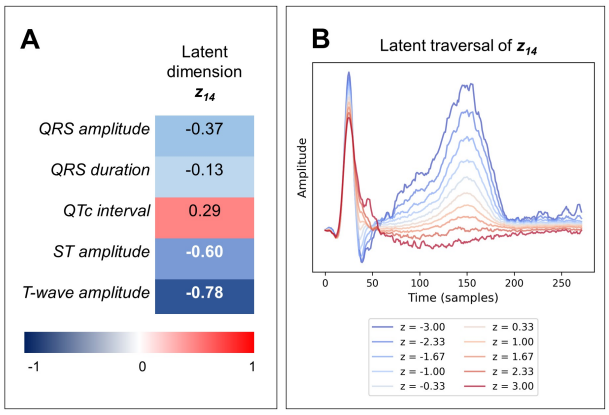

Figure 3. Sex-specific latent structure interpretability. (A) Correlation profile between latent dimension z_{14} and ECG parameters. (B) Latent traversal for z_{14} from -3 to $+3$.

late prominent axes of variation, they lack the mechanism to control in advance which specific dimension will store a given clinical factor[12]. Consequently, mapping dimensions to physiological traits must be done post-training, potentially leaving subtle components of sex-specific morphology entangled with other physiological sources of variation.

Nevertheless, the strong association of z_{14} with ventricular repolarization markers (ST-segment and T-wave amplitudes) is consistent with established sex-related differences in cardiac repolarization [1]. Future work will focus on extending the current framework towards conditional and semi-supervised approaches as Conditional Variational Autoencoders (CVAEs) [13]. By explicitly conditioning the latent space, CVAEs could enforce a strict mathematical separation of distributed traits, preventing sex-specific information from leaking into other dimensions. Crucially, this framework opens the door to incorporating underrepresented clinical covariates in female cardiovascular health, such as menopausal status or hormonal fluctuations, which are known to modulate ventricular repolarization, specifically the QT_c interval [2]. Refining the latent space to capture these female-specific physiological transitions will significantly advance the development of personalized, sex-informed digital biomarkers for automated CV risk assessment.

In conclusion, our findings suggest that an unsupervised FactorVAE successfully compresses single-lead ECG signals into 11 active LDs while capturing sex-specific variations within dimension z_{14}. By automatically disentangling this sex-specific physiological variability without explicit supervision, this framework provides a way for cleaner, unbiased electrocardiographic analysis and further risk stratification.

Acknowledgments

This work was supported by project PID2023-148975OB-I00, funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033). PBM acknowledges the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR-203330); a delivery partnership of Barts Health NHS Trust, QMUL, St George's University Hospitals NHS Foundation Trust and St George's University of London.

References

[1] Prajapati C, et al. Sex differences in heart: from basics to clinics. Eur J Med Res 2022;27(241).

[2] Linde C, et al. Sex differences in cardiac arrhythmia: a consensus document of the European Heart Rhythm association, endorsed by the Heart Rhythm Society and Asia Pacific Heart Rhythm Society. EP 2018;20(10):1565–1565ao.

[3] Wang X, et al. Sex hormones and reproductive factors with cardiac arrhythmia and ECG indices: a mendelian randomization study. BMC Cardiovasc Disord 2024;24(659).

[4] Kuznetsov VV, et al. Interpretable feature generation in ECG using a variational autoencoder. Front Genet 2021; 12:638191.

[5] Kingma DP, Welling M. Auto-Encoding Variational Bayes, arXiv preprint arXiv:1312.6114, 2013.

[6] Burgess CP, et al. Understanding disentangling in β-VAE, arXiv preprint arXiv: 1804.03599, 2018.

[7] Kim H, Mnih A. Disentangling by Factorising. Proceedings of the 35th ICML 2018;PMLR 80:2649–2658.

[8] Kapsecker M, Möller MC, Jonas SM. Disentangled representational learning for anomaly detection in single-lead electrocardiogram signals using variational autoencoder. Comp in Biol and Med 2025;184:109422.

[9] Sudlow C, et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases. PLoS Med 2015;12(3):e1001779.

[10] Madrid J, et al. Unsupervised clustering of single-lead electrocardiograms associates with prevalent and incident heart failure in coronary artery disease. EHJ Digital Health 2025; 6:435–446.

[11] Patlatzoglou K, et al. The cost of explainability in artificial intelligence-enhanced electrocardiogram models. npj Digit Med 2025;8(747).

[12] Almudévar A, Mariotte T, Ortega A, Tahon M. Unsupervised multiple domain translation through controlled disentanglement in variational autoencoder. In ICASSP. 2024; 7010–7014.

[13] Sohn K, Lee H, Yan X. Learning Structured Output Representation using Deep Conditional Generative Models. Advances in Neural Information Processing Systems 28, NeurIPS Proceedings, 2015.

Address for correspondence:

Ángela Hernández
C. Mariano Esquillor s/n, D-6.01.1B, 50018, Zaragoza, Spain
angela.hernandezm@unizar.es