ECG Morphology Representational Learning and Decomposition Via An Adaptive Legendre Mixture Model Variational Autoencoder

Kaveh Samiee

Doublepoint, Helsinki, Finland

Abstract

The representational learning of ECG wave complexes has emerged as a powerful approach for capturing clinically salient features from cardiac electrical activity while enabling sub-Nyquist data compression. The topology of the latent space plays a crucial role in maintaining fidelity and preserving the morphological characteristics of reconstructed waveforms. This paper proposes an Adaptive Legendre Mixture Model Variational Autoencoder (ALMM-VAE) for structured and interpretable ECG representation learning. By employing Legendre polynomial-based basis functions in a mixture model framework within the latent space, ALMM-VAE enables orthogonal, efficient, and compact representation of ECG signals. Experimental evaluations on the MIT-BIH Arrhythmia dataset demonstrate that ALMM-VAE achieves superior performance in waveform reconstruction and latent space interpretability compared with traditional Gaussian VAEs.

1. Introduction

Electrocardiography (ECG) provides essential insights into cardiac electrophysiology. With the proliferation of mobile health technologies, there is a growing need for robust representation learning frameworks capable of compressing ECG signals without sacrificing clinical fidelity. Deep generative models, particularly Variational Autoencoders (VAEs), have demonstrated promise in unsupervised learning of ECG representations. However, conventional VAEs, which assume Gaussian priors in the latent space, often struggle to encode structured and interpretable features critical for downstream clinical tasks.

In work introduces an adaptive Legendre Mixture Model Variational Autoencoder (ALMM-VAE), which structures its latent space using orthogonal Legendre polynomial basis functions. By adaptively weighting these components, the model captures detailed ECG morphology and enhances reconstruction of subtle waveform variations, while maintaining interpretability for diverse cardiac rhythms.

2. Related Work

Autoencoders and VAEs have been widely applied to ECG signal processing for dimensionality reduction and denoising tasks. We first present the mathematical principles of these architectures and identify the main limitations addressed by our proposed approach.

2.1. Conventional Autoencoders

Consider a dataset of ECG recordings represented as

$$\mathcal{D} \triangleq \{\mathbf{x_i}\}_{\mathbf{i=1}}^{\mathbf{M}},\ \mathbf{x_i} \in \mathbb{R}^{\mathrm{n}},\ \mathrm{n} \in \mathbb{N}, \tag{1}$$

where each $\mathbf{x_i}$ represents a discretized ECG segment of length n samples, extracted from long-term multi-channel recordings and sampled from the population distribution $p(\mathbf{x})$ of cardiac electrical activity. The main objective is to learn a compact, yet discriminative representation of each observation $\mathbf{x_i}$ that retains clinically relevant morphological patterns essential for diagnostic analysis. An autoencoder learns to compress and reconstruct data through two neural networks: an encoder $f_{\boldsymbol{\phi}} : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{q}}$ that maps input signals into a lower-dimensional latent code, and a decoder $g_{\boldsymbol{\theta}} : \mathbb{R}^{\mathrm{q}} \to \mathbb{R}^{\mathrm{n}}$ that reconstructs the original signal, where $q < n$ ensures effective compression.

The autoencoder parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are optimized by minimizing the mean squared reconstruction error (MSE):

$$\mathcal{L}_{\mathrm{AE}} = \mathbb{E}_{\mathbf{x}\sim\mathbf{p}(\mathbf{x})}\Big[\big\|\mathbf{x} - \mathbf{g}_{\boldsymbol{\theta}}\big(\mathbf{f}_{\boldsymbol{\phi}}(\mathbf{x})\big)\big\|^{\mathbf{2}}\Big]. \tag{2}$$

Despite the effectiveness of autoencoders in compressing and reconstructing input data, their generalization power is limited by two fundamental constraints. First, classical autoencoders are prone to overfitting and exhibit poor generalization performance, particularly when trained on limited datasets [1, 2]. Second, their deterministic latent spaces lack probabilistic structure, resulting in interpolation artifacts when decoding linear combinations of latent codes due to the non-smooth and non-convex nature of the learned manifold [3, 4].

2.2. Variational Autoencoders

VAEs [5] address this limitation by introducing probabilistic encoding. Instead of deterministic mappings, the encoder outputs parameters of a posterior distribution over the latent space. Specifically, the encoder produces the mean $\boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathbf{x})$ and variance $\boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\mathbf{x})$ vectors that parameterize a Gaussian posterior:

$$q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}) = \mathcal{N}\big(\mathbf{z}; \boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathbf{x}), \boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{\phi}}(\mathbf{x})\mathbf{I}\big). \tag{3}$$

This probabilistic formulation enables the VAE to capture uncertainty and learn a structured latent space. The training objective maximizes the evidence lower bound (ELBO), which can be written as:

$$\begin{aligned}\mathcal{L}_{\text{VAE}} = \mathbb{E}_{\mathbf{q}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\big[\log \mathbf{p}_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z})\big] \\ - D_{\text{KL}}\big(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}) \,\|\, \mathbf{p}(\mathbf{z})\big).\end{aligned} \tag{4}$$

The first term encourages faithful reconstruction, while the second term, known as the Kullback-Leibler (KL) divergence, regularizes the posterior to remain close to the prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. This regularization ensures that the latent space maintains a well-structured topology suitable for sampling and interpolation.

During inference, the VAE samples from the learned posterior distribution:

$$\mathbf{z} \sim \mathbf{q}_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}), \tag{5}$$

using the reparameterization trick to enable backpropagation. The sampled latent code is then decoded to produce the reconstruction:

$$\hat{\mathbf{x}} = g_{\boldsymbol{\theta}}(\mathbf{z}) \ \in \ \mathbb{R}^{\mathrm{n}}. \tag{6}$$

Critically, all latent operations occur in the compressed space $\mathbb{R}^{\mathrm{q}}$, while both input and reconstructed signals reside in the original space $\mathbb{R}^{\mathrm{n}}$. This dimensional separation is essential for computational efficiency and interpretability.

3. Legendre Mixture Model Transformation in VAE

Standard VAEs with Gaussian priors often struggle to capture complex, multi-modal distributions present in ECG data. Different types of arrhythmias may form distinct clusters in the latent space, but a single Gaussian prior cannot adequately model such structured variations. Furthermore, learned representations may lack clinical interpretability, limiting their utility for diagnostic applications. Gaussian Mixture VAEs (GM-VAEs) [6] attempt to address this by replacing the standard Gaussian prior with a mixture model:

$$p(\mathbf{z}) = \sum_{\mathbf{k=1}}^{\mathbf{K}} \pi_{\mathbf{k}} \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_{\mathbf{k}}, \boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{k}}\mathbf{I}), \tag{7}$$

where π_k are mixture weights. While this increases modeling flexibility, it often lacks the structured interpretability crucial for clinical applications. These limitations motivate our proposed ALMM-VAE, which employs orthogonal Legendre polynomial basis functions to create a more structured and interpretable latent space specifically suited for ECG morphology analysis.

Legendre polynomials $\{P_n(t)\}_{n=0}^{\infty}$ form a complete orthogonal basis functions on the interval $[-1, 1]$, satisfying the fundamental orthogonality relation:

$$\int_{-1}^{1} P_m(t) P_n(t)\, dt = \frac{2}{2n+1}\delta_{mn}, \tag{8}$$

where δ_{mn} is the Kronecker delta. The first few Legendre polynomials are defined as:

$$\begin{aligned}&P_0(t) = 1, \quad P_1(t) = t, \quad P_2(t) = \frac{1}{2}(3t^2 - 1), \\ &P_3(t) = \frac{1}{2}(5t^3 - 3t), \quad P_4(t) = \frac{1}{8}(35t^4 - 30t^2 + 3).\end{aligned} \tag{9}$$

For ECG signal analysis, Legendre polynomials offer several advantages over conventional basis functions: 1) orthogonality, ensures minimal correlation between basis components, leading to compact representations; 2) finite support, natural fit for ECG segments with bounded temporal duration; 3) morphological preservation, low-order polynomials capture global trends while higher-order terms encode local morphological variations; 4) analytical tractability, closed-form expressions facilitate efficient computation and interpretability.

In the ALMM-VAE framework, we maintain the standard Gaussian prior for stable training:

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I}), \tag{10}$$

while introducing structured mixture modeling through a deterministic transformation applied to latent space $\mathbf{z}$. During generation, each mixture component $k = 1, \ldots, K$ represents a truncated Legendre polynomial expansion:

$$f_k(\tilde{\mathbf{z}}) = \sum_{n=0}^{D_k} \mathbf{a_{k,n}} \mathbf{P_n}(\tilde{\mathbf{z}}), \tag{11}$$

where $D_k \in \{0, 1, \ldots, D_{\max}\}$ is the learnable polynomial degree for component k, $\mathbf{a_{k,n}} \in \mathbb{R}^{\mathbb{D}}$ are learnable Legendre coefficients defining the contribution of each Legendre polynomial $P_n(\cdot)$ across all latent dimensions, and $\tilde{\mathbf{z}}$ denotes the normalized latent space obtained using the normalization transform:

$$\tilde{\mathbf{z}} = 2\frac{\mathbf{z} - \mathbf{z}_{\min}}{\mathbf{z}_{\max} - \mathbf{z}_{\min}} - 1, \tag{12}$$

where $\mathbf{z}_{\min}, \mathbf{z}_{\max}$ denote the element-wise minimum and maximum of $\mathbf{z}$. This normalization over the latent space ensures that all operations occur within the canonical Legendre domain $[-1, 1]$,

The structured latent representation is computed as:

$$\mathbf{h}(\mathbf{z}) = \sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{w_k} \cdot \mathbf{f_k}(\tilde{\mathbf{z}}), \tag{13}$$

where $w_k = \text{softmax}(\alpha_k)$ are the normalized mixture weights derived from the learnable parameters $\boldsymbol{\alpha} \in \mathbb{R}^{\mathbb{K}}$ and satisfy the normalization constraint $\sum_{k=1}^{K} w_k = 1$, ensuring proper mixture combination.

The ALMM-VAE employs a dual-mode architecture: during training, the decoder operates on original latent codes $\mathbf{z}$ for stable reconstruction, while during generation, it utilizes the transformed representation $\mathbf{h}(\mathbf{z})$ for structured sampling. The training objective extends the standard ELBO with regularization on Legendre parameters:

$$\begin{aligned}\mathcal{L}_{\text{ALMM-VAE}} = \mathbb{E}_{\mathrm{q}_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\big[\log \mathrm{p}_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{h}(\mathbf{z}))\big] \\ - D_{\mathrm{KL}}\big(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}) \,\|\, \mathbf{p}(\mathbf{z})\big) + \mathbf{R}(\boldsymbol{\alpha}, \mathbf{a}),\end{aligned} \tag{14}$$

where the regularization term promotes coefficient sparsity and balanced mixture utilization:

$$\begin{aligned}R(\boldsymbol{\alpha}, \mathbf{a}) = \beta_{\mathbf{1}} \sum_{\mathbf{k=1}}^{\mathbf{K}} \sum_{\mathbf{n=0}}^{\mathbf{D}_{\max}} \|\mathbf{a}_{\mathbf{k,n}}\|_{\mathbf{2}}^{\mathbf{2}} \\ +\beta_2 \left(-\sum_{k=1}^{K} w_k \log(w_k + \epsilon) \right),\end{aligned} \tag{15}$$

with β_1 and β_2 controlling coefficient regularization and promoting mixture diversity through entropy maximization, respectively. The objective function in Eq. (14), maintains the computational simplicity of standard VAE in Eq. (4) while allows mixture components to automatically adjust their polynomial degree D_k to match the complexity of different ECG patterns.

4. Experimental Setup

4.1. Dataset

In this study, we used the Massachusetts Institute of Technology and Beth Israel Hospital (MIT-BIH) arrhythmia database [7]. MIT-BIH contains 48 records of 30-minute duration, each comprising two-channel ambulatory ECG data obtained from 47 subjects. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10-millivolt range, providing approximately 110,000 expert-annotated heartbeats across all records. In this study, we exclusively utilize the Lead II channel of each record due to its optimal representation of ventricular depolarization. ECG signals are first filtered with a band-pass filter (0.65-40 Hz) to remove baseline drift and high-frequency noise. Then, using the provided annotations, ECG segments are extracted by centering 300-millisecond square windows around each R-peak. The class label is then associated with each segment according to the five-class AAMI standard classification system.

4.2. Results

To benchmark the effectiveness of the proposed ALMM-VAE in the analysis of cardiac arrhythmia, a baseline VAE architecture is implemented with a latent dimension of $q = 10$ and symmetric encoder-decoder networks. Using the baseline VAE network, we then create two variants for ALMM-VAE and GMM-VAE models, with three mixture components for each. Furthermore, β_1 and β_2 in Eq. (15) are set to 0.001 and 0.01, respectively. The ECG segment dataset is divided into training and test sets with 60/40 split ratio. Reconstruction quality is evaluated using two complementary metrics: MSE, as defined in Eq. (2), for measuring absolute reconstruction fidelity, and cosine similarity [8] to assess directional alignment of the ECG morphology. The cosine similarity between original and reconstructed segments is computed as:

$$\cos(\theta_i) = \frac{\mathbf{x_i} \cdot \hat{\mathbf{x}}_\mathbf{i}}{||\mathbf{x_i}|| \times ||\hat{\mathbf{x}}_\mathbf{i}||}, \tag{16}$$

a high cosine similarity indicates morphological similarity between original segment $\mathbf{x_i}$ and its reconstruction $\hat{\mathbf{x}}_i$. Table 1 presents the reconstruction performance metrics averaged over 10 repetitions. The consistently high cosine similarity values obtained for both models indicate excellent preservation of ECG signal morphology, suggesting that the essential characteristics of cardiac waveforms are maintained during the encoding-decoding process. The minimal difference between training and test performance demonstrates good generalization capability without significant overfitting. Figure 1, shows visual comparison between original ECG segments and their reconstruction obtained by ALMM-VAE and GMM-VAE. To evaluate the discrimination power of latent space features obtained using each model, we leverage a binary classification scheme by mapping the five-class AAMI into ventricular ectopic beats versus all other classes. Classification performance was then assessed using the following metrics: overall accuracy (Acc), true positive rate or sensitivity (TPR), and positive predictive value or precision (PPV). Table 2 presents the classification performance results. ALMM-VAE demonstrates superior classification performance across all evaluated metrics, notably obtains higher sensitivity rate (88.79% vs 86.16%) and precision

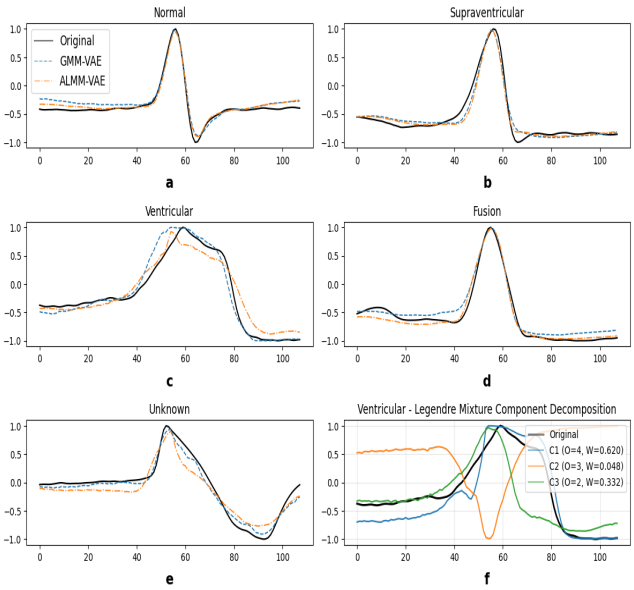


Figure 1. Visual comparison of original signals and their reconstructions using GMM-VAE and ALMM-VAE architectures across five ECG beat categories: a) normal beat, b) supra-ventricular beat, c) ventricular beat, d) fusion beat, and e) unknown beat. Panel f) illustrates the Legendre polynomial component orders and their corresponding mixture weights for a ventricular beat example.

(94.31% vs 93.02%), indicating a higher discriminative capability for detection of cardiac rhythm abnormalities.

Table 1. Average reconstruction performance metrics over 10 repetitions

Metric	**Set**	**ALMM-VAE**		**GMM-VAE**	
		Mean	**Std**	**Mean**	**Std**
Cosine Sim.	Train	0.9674	0.0504	**0.9695**	**0.0440**
	Test	0.9671	0.0511	**0.9694**	**0.0460**
MSE	Train	0.0224	0.0234	**0.0216**	**0.0205**
	Test	0.0225	0.0239	**0.0217**	**0.0211**

The experimental results show that the lower reconstruction error does not necessarily correlate with optimal domain-specific performance. The latent representations learned by ALMM-VAE, despite resulting in slightly higher reconstruction error, capture more discriminative clinically relevant features suitable for arrhythmia classification.

Table 2. Average classification performance metrics over 10 repetitions

Model	Acc%	TPR%	PPV%
ALMM-VAE	**98.91**	**88.79**	**94.31**
GMM-VAE	98.67	86.16	93.02

5. Conclusion

We introduce the ALMM-VAE, a novel VAE that employs Legendre polynomial mixtures within the latent space for ECG signal analysis. The experimental results show that ALMM-VAE produces compact, orthogonal, and interpretable latent representations. The main advantage of the proposed Legendre mixture model transformation arises from the orthogonal nature of Legendre polynomials, which substantially decreases correlation among mixture components. Moreover, our regularization strategy permits mixture components to optimize their polynomial degrees, facilitating the discovery and representation of unique cardiac patterns. Future work will focus on multi-lead ECG modeling and real-time wearable applications.

References

[1] Cremer C, Li X, Duvenaud D. Inference suboptimality in variational autoencoders. In Proceedings of the International Conference on Learning Representations. ICLR, 2018; 1–14.

[2] Steck H. Autoencoders that don't overfit towards the identity. In Advances in Neural Information Processing Systems, volume 33. Curran Associates, Inc., 2020; 19598–19608.

[3] Oring A, Yakhini Z, Hel-Or Y. Autoencoder image interpolation by shaping the latent space. In Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research. PMLR, 2021; 7326–7336.

[4] Xiao TZ, Zenn J, Bamler R. A note on generalization in variational autoencoders: How effective is synthetic data & overparameterization? Transactions on Machine Learning Research 2024;1(1):1–11.

[5] Kingma DP, Welling M. Auto-encoding variational bayes. In Proceedings of the International Conference on Learning Representations. ICLR, 2014; 1–13.

[6] Dilokthanakul N, Mediano PAM, Garnelo M, Lee MCH, Salimbeni H, Arulkumaran K, Shanahan M. Deep unsupervised clustering with gaussian mixture variational autoencoders. In Proceedings of the International Conference on Learning Representations. ICLR, 2017; 1–12.

[7] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine 2001;20(3):45–50.

[8] Chou EP, Hsu SM. Cosine similarity as a sample size-free measure to quantify phase clustering within a single neurophysiological signal. Journal of Neuroscience Methods 2018;295:1–10. Application of cosine similarity in biomedical signal analysis.

Address for correspondence:

Kaveh Samiee
Pursimiehenkatu 26 C, 00150 Helsinki, Finland.
kaveh@doublepoint.com