Physiology-Informed Wavelet State Space Modeling for Interpretable ECG Arrhythmia Classification

Kaveh Samiee
Independent Researcher, Helsinki, Finland

Abstract

Deep learning achieves strong electrocardiogram (ECG) arrhythmia classification but often provides limited insight into the morphology driving its predictions. We present Wavelet SSM, a physiology-informed state space model (SSM) for interpretable single-beat classification. A frozen transition matrix is constructed from sampled wavelet atoms under a QRS-centered measure, making one state axis explicitly multiscale, while input-dependent write and read vectors retain content selectivity. A convolutional stem reduces the sequence length before three residual Wavelet SSM blocks, and a learned scale-weighting vector contributes the final wavelet state to the classifier. In preliminary five-class Association for the Advancement of Medical Instrumentation (AAMI) experiments on the Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH) Arrhythmia Database, the Daubechies db3 variant attains 93.23% *F1 under the implemented beat-level split, remaining competitive with results reported for the same five-class MIT-BIH task. The model also exposes class-conditioned state gradients over known wavelet scales, providing an inspectable multiscale representation for beat labeling.*

1. Introduction

Interpretability, limited access to annotated clinical data, and resource-efficient inference remain fundamental requirements for clinically useful artificial intelligence. In cardiac monitoring, reliable beat-by-beat electrocardiogram (ECG) analysis supports both screening and continuous monitoring. Automated classification has progressed from handcrafted morphology and rhythm features to end-to-end networks with expert-level performance on selected tasks [1]. Their latent representations, however, are rarely aligned with the morphological vocabulary used by clinicians, making individual decisions difficult to inspect.

The wavelet transform provides a natural representation of ECG morphology. Its joint time-frequency localization captures the multiscale structure of the P wave, QRS complex, and T wave, and has long supported ECG delineation and feature extraction [2]. Classical wavelet pipelines, however, rely on manually selected coefficients and shallow classifiers. State space models (SSMs) offer a learnable alternative. High-order Polynomial Projection Operators (HiPPO) introduced continuous-time dynamics that project signal histories onto orthogonal polynomial bases [3]; S4 made these structured dynamics practical for long-sequence modeling [4], and Mamba added input-dependent selectivity with linear sequence complexity [5]. Selective SSMs have subsequently been adapted to ECG using multibranch and bidirectional architectures for arrhythmia and 12-lead classification [6], including transparency-oriented variants [7].

The HiPPO projection view has also been generalized beyond polynomial bases to arbitrary orthogonal bases [8]. SaFARi extended this construction numerically to general frames, including wavelets [9], while recent WaveSSM work represented nonstationary and physiological signals using localized, multiscale latent states [10]. These studies establish the general wavelet and frame-based SSM construction, but do not specialize its scale selection, projection measure, or interpretation to cardiac physiology.

This work focuses on the intersection of wavelet representations and selective SSMs for ECG. Its contributions are: (i) a cardiac-conditioned transition matrix constructed from sampled wavelet atoms and a QRS-centered measure, then frozen within a selective SSM recurrence; (ii) a compact architecture combining a convolutional stem, input-dependent state write/read vectors, and an explicit scale-state readout; and (iii) an empirical comparison of four wavelet families for five-class Association for the Advancement of Medical Instrumentation (AAMI) beat classification, together with class-conditioned scale attribution.

2. Methods

2.1. Beat representation and feature lifting

Let $x^{(\ell)}[n] \in \mathbb{R}$ denote a discrete ECG signal from a single lead ℓ, sampled at frequency f_s. After QRS localization from reference annotations, let $\{r_i\}$ be the R-peak indices. For even L, the ith beat is the centered segment

$$u_i[m] = x^{(\ell)}[r_i + m], \quad m = -L/2, \ldots, L/2 - 1, \quad (1)$$

Corresponding author: Kaveh Samiee, kaveh.samiee@gmail.com

with class label $y_i \in \mathcal{Y}$.

A two-stage convolutional stem lifts the scalar beat to a D-dimensional feature sequence. Each stage consists of a same-padded one-dimensional convolution (kernels of length 7 then 5), batch normalization (BN), ReLU activation, and max pooling by two. A pointwise linear projection then maps the final channels to D features, so the stem sends $\mathbb{R}^L$ to $\mathbb{R}^{T\times D}$ with $T = L/4$. Fig. 1 summarizes the pipeline.

2.2. Wavelet-structured transition

HiPPO constructs a linear SSM by projecting a signal history onto a prescribed basis [3]; the same idea has been extended to general orthogonal bases and frames, including wavelets [8,9]. We do not instantiate that ODE. Instead we freeze a discrete transition whose coordinates are indexed by a finite dictionary of R-centered wavelet dilations.

Let $\Omega = [-1, 1]$ and let ψ be a mother wavelet. With $(s_{\min}, s_{\max}) = (0.3, 3.0)$, the scales are

$$s_n = s_{\min}\big(s_{\max}/s_{\min}\big)^{(n-1)/(N-1)}. \tag{2}$$

The nth atom is $\phi_n(\tau) = s_n^{-1/2}\psi(\tau/s_n)$, where ψ is obtained from a cascade approximation (refinement level 8) and recentered at the origin; for biorthogonal families the analysis wavelet is used. Sampling on M=90 uniform points $\tau_j \in \Omega$ yields $\Phi_{j,n} = \phi_n(\tau_j)$, after which each column is ℓ_2-normalized. This grid is independent of the beat length L and of the stem length T. A QRS-centered weighting $w_j \propto \exp(-2\tau_j^2)$, normalized so that $\sum_j w_j = 1$, defines $\mathbf{W} = \mathrm{diag}(\mathbf{w_1}, \ldots, \mathbf{w_M})$. With $\mathbf{G_0} = \mathbf{\Phi}^\top \mathbf{W}\mathbf{\Phi}$ and ridge $\epsilon = 10^{-8}$,

$$\begin{aligned} \mathbf{A}_{\mathrm{wav}} &= -\mathrm{diag}(\mathbf{d}) + \mathbf{0.1}(\mathbf{G}_\epsilon^{-1}\mathbf{G_0} - \mathbf{I}), \\ d_n &= \exp\Big[-\tfrac{1}{2}\tfrac{n-1}{N-1}\Big], \end{aligned} \tag{3}$$

where $\mathbf{G}_\epsilon = \mathbf{G_0} + \epsilon\mathbf{I}$. Equivalently, the second term is the scaled inverse-Gram damping $-0.1\epsilon(\mathbf{G_0} + \epsilon\mathbf{I})^{-1}$. If the spectral radius $\rho(\mathbf{A}_{\mathrm{wav}})$ exceeds 0.99, the matrix is rescaled so that $\rho = 0.95$. The resulting matrix is used directly as a discrete map (no zero-order-hold step) and is frozen during training. Negative diagonal entries make the discrete recurrence oscillatory after rescaling. The state axis remains indexed by the prescribed scales s_n, while write and read vectors are learned; this is a structured prior, not a HiPPO projection of the ECG.

2.3. Selective Wavelet SSM block

Each of the three residual blocks is preceded by root mean square normalization (RMSNorm) and stores a frozen copy of the same $\mathbf{A}_{\mathrm{wav}}$. For block input $\mathbf{x_t} \in \mathbb{R}^{\mathbf{D}}$, a linear map is split into a content branch and a gate branch. A causal depthwise convolution (kernel length 4) and Sigmoid Linear Unit (SiLU) activation produce $\mathbf{v_t} \in \mathbb{R}^{\mathbf{D_e}}$ with $D_e = 2D$. A linear selection map of $\mathbf{v_t}$ yields a single pair $\mathbf{b_t}, \mathbf{c_t} \in \mathbb{R}^{\mathbf{N}}$, shared across channels. The state is initialized at $\mathbf{h_{0,q}} = \mathbf{0}$ and reset per beat. For expanded channel q,

$$\mathbf{h_{t,q}} = \mathbf{A}_{\mathrm{wav}}\mathbf{h_{t-1,q}} + \mathbf{b_t}\,\mathbf{v_{t,q}}, \tag{4}$$

$$o_{t,q} = \mathbf{c_t}^\top \mathbf{h_{t,q}} + \delta_{\mathbf{q}}\mathbf{v_{t,q}}. \tag{5}$$

The output is gated by SiLU of the gate branch, projected back to D, and added to the block input. This retains Mamba-style input-dependent write/read vectors, but the recurrence uses the frozen discrete matrix directly: it does not learn a step size or apply zero-order-hold discretization.

2.4. Scale readout and objective

After the final block, temporal mean pooling gives $\mathbf{p} \in \mathbb{R}^{\mathbf{D}}$. The last-block state $\mathbf{H_T} \in \mathbb{R}^{\mathbf{D_e}\times\mathbf{N}}$ is averaged over expanded channels to give $\mathbf{h}^\star \in \mathbb{R}^{\mathbf{N}}$. A global learned scale vector $\boldsymbol{\theta} \in \mathbb{R}^N$ forms

$$\mathbf{r} = \mathbf{p} + \mathbf{W_h}\left[\sigma(\boldsymbol{\theta}) \odot \mathbf{h}^\star\right], \qquad \mathbf{z} = \mathbf{W_o}\,\mathrm{RMSNorm}(\mathbf{r}), \tag{6}$$

where σ is the sigmoid, $\mathbf{W_h} \in \mathbb{R}^{\mathbf{D}\times\mathbf{N}}$, and $\mathbf{W_o} \in \mathbb{R}^{|\mathcal{Y}|\times\mathcal{D}}$. Dropout with rate 0.2 is applied after RMSNorm. The scale weights are shared across beats rather than generated from each input. A second linear head on the same normalized representation produces ventricular-versus-rest logits (class V versus $\{\mathrm{N}, \mathrm{S}, \mathrm{F}, \mathrm{Q}\}$), and the training loss is $\mathcal{L} = \mathcal{L}_{\bigtriangledown} + \mathit{1}.\bigtriangledown\mathcal{L}_{\mathrm{V/bin}}$.

2.5. Scale attribution

For target class k, gradients of the class logit z_k are taken with respect to the final-block state and averaged over expanded channels and beats of class k:

$$a_{k,n} = \frac{1}{|\mathcal{D}_\parallel|\,|\mathcal{D}_\urcorner|}\sum_{i\in\mathcal{D}_\parallel}\sum_{q=1}^{D_e}\left|\frac{\partial z_k^{(i)}}{\partial h_{T,q,n}^{(i)}}\right|. \tag{7}$$

Each class profile is row-normalized for visualization. The index n labels a prescribed wavelet scale, but $\mathbf{h_T}$ is a latent of the stem features after three residual blocks, not a continuous wavelet transform of the ECG; the gradient is a post hoc sensitivity of the class score to that state.

3. Datasets and Experiments

We use the Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH) Arrhythmia Database (48 recordings) sampled at $f_s = 360$ Hz [11]. Reference annotations provide the R-peak indices $\{r_i\}$ and are mapped to the five AAMI classes $\mathcal{Y} = \{\mathrm{N}, \mathrm{S}, \mathrm{V}, \mathrm{F}, \mathrm{Q}\}$ [12], namely normal or non-ectopic (N), supraventricular ectopic (S), ventricular ectopic (V), fusion of normal and ventricular beats (F), and paced or otherwise unclassifiable beats (Q). We set

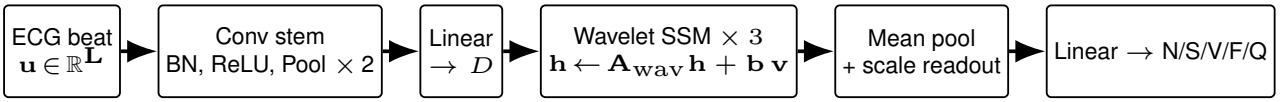

Figure 1: Simplified Wavelet SSM pipeline (residuals omitted). An ECG beat is lifted by a convolutional stem, processed by residual blocks with frozen $\mathbf{A}_{\mathrm{wav}}$, then mean-pooled; a learned scale readout mixes the final state into the classifier features.

Table 1: Five-class MIT-BIH beat classification performance.

Method	Acc%	PPV%	TPR%	F1%
Wavelet SSM, db2	98.29	95.20	90.17	92.62
Wavelet SSM, db3	98.48	**95.72**	90.86	**93.23**
Wavelet SSM, bior3.3	98.31	94.80	90.07	92.37
Wavelet SSM, db4	97.97	93.44	90.05	91.71
S6 (Mamba)	98.31	94.76	89.72	92.17
S4 (HiPPO)	97.95	93.72	89.52	91.57
Kiranyaz et al. [13]	95.90	84.20	88.80	86.40
Zhai et al. [14]	96.80	87.90	92.00	89.90
Shaker et al. [15]	**98.70**	90.10	**95.90**	92.90
Li et al. [16]	92.00	62.80	93.30	75.10
Zhou et al. [17]	97.90	90.80	89.70	90.20

$L = 180$ and extract a 500 ms window centered on each annotated R peak from every available lead, treating each lead window as an individual example. Each window is standardized by its own mean and standard deviation, then rescaled to $[0, 1]$. The archived experiments use a seeded beat-level random split of $60/10/30\%$ for training, validation, and test. Consequently, the present numbers measure beat-level interpolation, not generalization to unseen patients.

The stem compresses each beat to a length-T=45 sequence that is then processed by three residual Wavelet SSM blocks with shared width (D=64, expanded channels D_e=128, state order N=32). Mother-wavelet choice is treated as an experimental factor: db2, db3, db4, and bior3.3 are compared under matched capacity and training protocol (20 epochs, batch size 128, Adam with cosine learning-rate decay, seed 42; best validation macro F1 checkpoint).

We report accuracy together with macro-averaged positive predictive value (PPV), true positive rate (TPR), and their harmonic mean as F1. The S6 comparison uses the same convolutional dimensions and number of blocks, but retains learned step-size discretization and selective scan; it is therefore a backbone comparison rather than an ablation of $\mathbf{A}$ alone. Representative published methods evaluated on the same five-class MIT-BIH task are included for context using the metrics reported in their respective studies.

4. Results

Table 1 compares the Wavelet SSM variants with SSM baselines and representative results reported for the same five-class MIT-BIH task. Among the wavelet variants of the proposed method, db3 attains the highest F1 of 93.23%, PPV of 95.72%, and TPR of 90.86%, followed by db2 and bior3.3. Relative to the matched S6 and S4 backbones and to the cited literature, db3 also records the highest F1 and PPV in the table, exceeding S6 and S4 by 1.06 and 1.66 F1 points, respectively. Shaker et al. [15] report the highest accuracy and TPR. These literature comparisons remain contextual, since lead selection, data partitioning, preprocessing, and metric implementations can differ between studies.

Fig. 2 demonstrates that the fixed scale axis can be interrogated through class-conditioned gradients; the class-averaged profiles vary across labels, with a shared mid-fine peak near scales 5–7. Fig. 3 shows a locked-split N/V pair (confidence 1.0; energy-duration width proxies $\sim$56 ms vs. $\sim$217 ms): the N beat concentrates gradient mass on finer/mid scales, whereas the V beat yields a broader, less fine-peaked profile. The analysis remains descriptive: gradients are post hoc, and no uncertainty estimate or hypothesis test yet supports a population fine-to-coarse ordering.

5. Conclusion and Future Work

Wavelet SSM combines a QRS-centered wavelet transition with input-dependent state write and read vectors for single-beat ECG classification. On the present MIT-BIH setup, the db3 variant is competitive with matched SSM baselines and with published five-class scores, while class-conditioned state gradients provide a direct view of the frozen scale axis. The current evaluation is intradatabase and beat-level, so patient-level generalization remains an open question. Future work will move to record-disjoint splits and assess whether the same wavelet prior and scale attributions transfer across patients and databases.

References

[1] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine 2019; 25(1):65–69.

[2] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989;11(7):674–693.

[3] Gu A, Dao T, Ermon S, Rudra A, Ré C. HiPPO: Recurrent memory with optimal polynomial projections. In Advances in Neural Information Processing Systems (NeurIPS). 2020; .

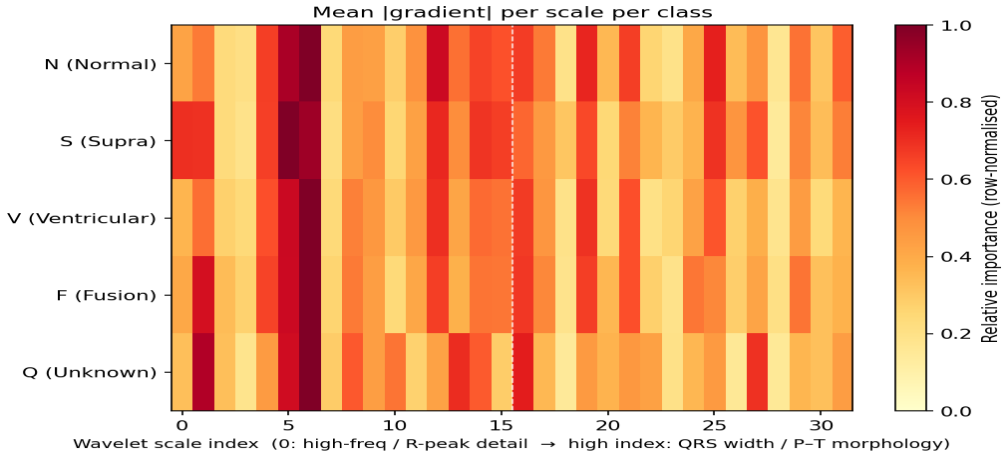

Figure 2: Class-conditioned wavelet-state attribution using db3 on the test set: row-normalized mean $|\partial z_k / \partial h_{T,\cdot,n}|$. Scale indices increase from fine to coarse.

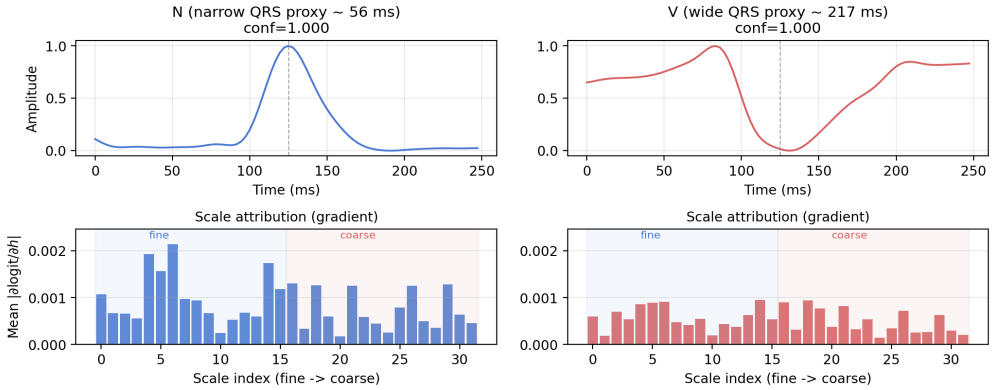

Figure 3: Locked-split Normal (N) vs. Ventricular (V) examples (Wavelet SSM, db3): R-peak–centered waveforms (top) and gradient attribution $|\partial \mathrm{logit} / \partial \mathrm{h}|$ over wavelet scales (fine→coarse, bottom).

[4] Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. In International Conference on Learning Representations (ICLR). 2022; .

[5] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv231200752 2023;.

[6] Jiang H, Mutahira H, Wei S, Muhammad MS. ECG-Mamba: Cardiac abnormality classification with non-uniform-mix augmentation on 12-lead ECGs. IEEE Journal of Translational Engineering in Health and Medicine 2025; 13:461–470.

[7] MambaCapsule authors. MambaCapsule: Towards transparent cardiac disease diagnosis with electrocardiography using mamba capsule network. arXiv preprint arXiv240720893 2024;.

[8] Gu A, Johnson I, Timalsina A, Rudra A, Ré C. How to train your HiPPO: State space models with generalized orthogonal basis projections. In International Conference on Learning Representations (ICLR). 2023; .

[9] Babaei H, White M, Alemohammad S, Baraniuk RG. SaFARi: State-space models for frame-agnostic representation. arXiv preprint arXiv250508977 2025;.

[10] WaveSSM authors. WaveSSM: Multiscale state-space models for non-stationary signal attention. arXiv preprint arXiv260222266 2026;.

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

[12] ANSI/AAMI. Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. ANSI/AAMI EC57:2012, 2012.

[13] Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-d convolutional neural networks. IEEE Transactions on Biomedical Engineering 2016;63(3):664–675.

[14] Zhai X, et al. Ecg arrhythmia beat classification with deep neural networks. VERIFY venue 2020;.

[15] Shaker AM, et al. Generalization of convolutional neural networks for ecg classification using generative adversarial networks. VERIFY venue 2021;.

[16] Li, et al. Deep learning for ecg arrhythmia classification. VERIFY venue 2021;.

[17] Zhou, et al. Ecg heartbeat classification with deep neural networks. VERIFY venue 2022;.