Supplementary Information

Interpretable Classification of Time Series Using Euler Characteristic Surfaces

Buddha Nath Sharma, Vishal Mandal, Salam Rabindrajit Luwang, Md. Nurujjaman,

Atish J. Mitra, Sushovan Majhi

Contents

1 Geometric Constructions: Voronoi Cells, Nerve, and Alpha Complex

This section provides geometric illustrations and formal definitions of the Voronoi diagram, the nerve (Čech complex), and the Alpha complex—the building blocks of our topological pipeline.

Formal definitions. Given a collection of subsets $\mathcal{U} = \{U_\alpha\}_{\alpha\in\Lambda}$ of a topological space, the *nerve* of $\mathcal{U}$ is the simplicial complex $N(\mathcal{U})$ whose vertex set is Λ, and where a subset $\{\alpha_0, \alpha_1, \ldots, \alpha_k\} \subseteq \Lambda$ spans a k-simplex if and only if $U_{\alpha_0} \cap U_{\alpha_1} \cap \cdots \cap U_{\alpha_k} \neq \emptyset$.

For a dataset $P = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$ and closed ball $B(x_i, r)$, the *Čech complex* at scale r is:

$$C(P, r) = \left\{ \sigma \subset P : \bigcap_{x_i\in\sigma} B(x_i, r) \neq \emptyset \right\}.$$

The *Voronoi cell* for $x_i \in P$ is $V(x_i) = \{x \in \mathbb{R}^d \mid d(x, x_i) \leq d(x, x_j),\ \forall x_j \in P\}$. The *Alpha complex* at filtration scale r is the nerve of the collection $\{V(x_i) \cap B(x_i, r)\}_{x_i\in P}$ [1].

Illustration. We consider six points forming a regular hexagon:

$$S = \{(0, 2), (2, 1), (2, -1), (0, -2), (-2, -1), (-2, 1)\}.$$

Figure S1 shows the Voronoi cells, the Čech complex (nerve), and the Alpha complex for this dataset. In Fig S1 (a), green dotted lines divide the space into six Voronoi regions. In Fig S1 (b), the Čech complex at a fixed radius includes all edges and triangles wherever corresponding balls mutually intersect. In Fig S1 (c), the Alpha complex restricts the Čech complex to simplices consistent with the Voronoi decomposition, yielding a sparser complex that still captures the essential topology.

2 Classification Flowcharts

Single-Feature Threshold Classifier

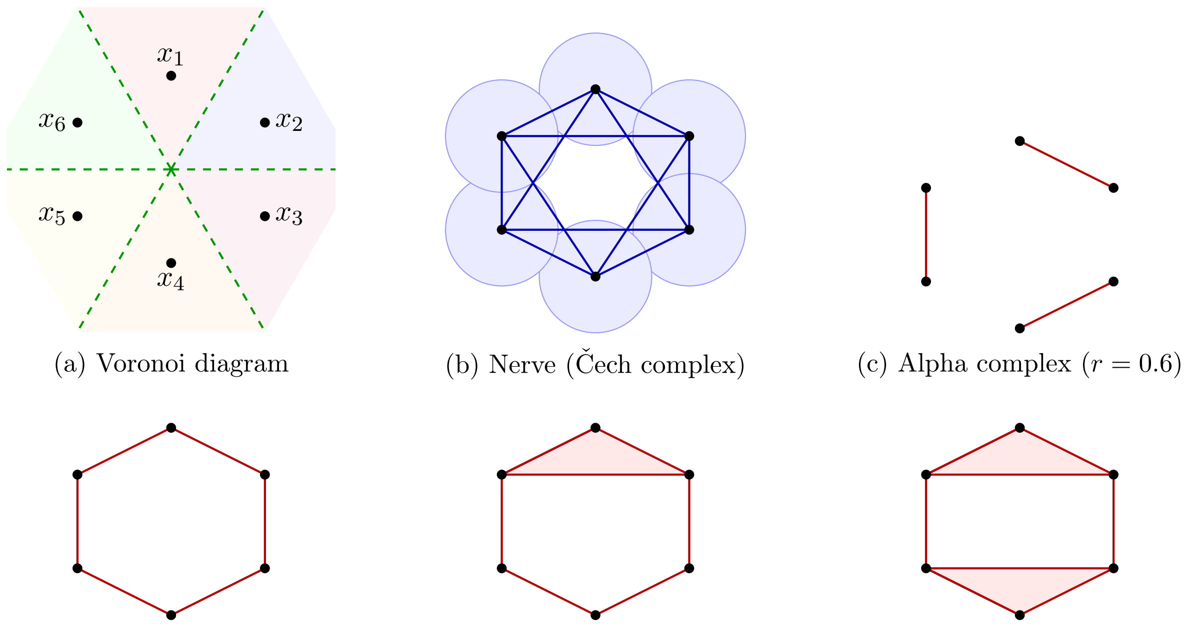
The flowchart in Fig. S2 illustrates the complete workflow of the single-feature ECS-based classification framework described in the Methods section of the main manuscript.

AdaBoost Ensemble Classifier

The flowchart in Fig. S3 illustrates the complete workflow of the ECS-based AdaBoost classification framework.

3 Proof of Stability Results

This section provides full proof of Lemma S1 stated in the main manuscript (therein referred to as Lemma 1). We first recall the relevant background from algebraic topology.



(a) Voronoi diagram (b) Nerve (Čech complex) (c) Alpha complex ($r = 0.6$)

(d) Alpha complex ($r = 1.0$) (e) Alpha complex ($r = 1.4$) (f) Alpha complex ($r = 1.8$)

Figure S1: Illustration of the Voronoi diagram, Nerve (Čech complex), and Alpha complexes for increasing filtration scales r for the hexagonal point set S. The sequence of Alpha-complex panels demonstrates how simplices are progressively added as the filtration scale increases.

Betti numbers and Euler characteristic. Using homology of a simplicial complex $\mathcal{K}$, an equivalent definition of its Euler characteristic is:

$$\chi(\mathcal{K}) = \sum_{i=0}^{\dim(\mathcal{K})} (-1)^i \, \beta_i(\mathcal{K}), \tag{S1}$$

where β_i is the i-th Betti number of $\mathcal{K}$ [2].

Persistence diagrams and Wasserstein distance. The *persistence diagram* $\mathrm{PD}_n(\mathcal{K})$ of a filtration $\mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_s$ and homological degree n is the multiset of points $(a_i, a_j) \in \overline{\mathbb{R}}^2$ with

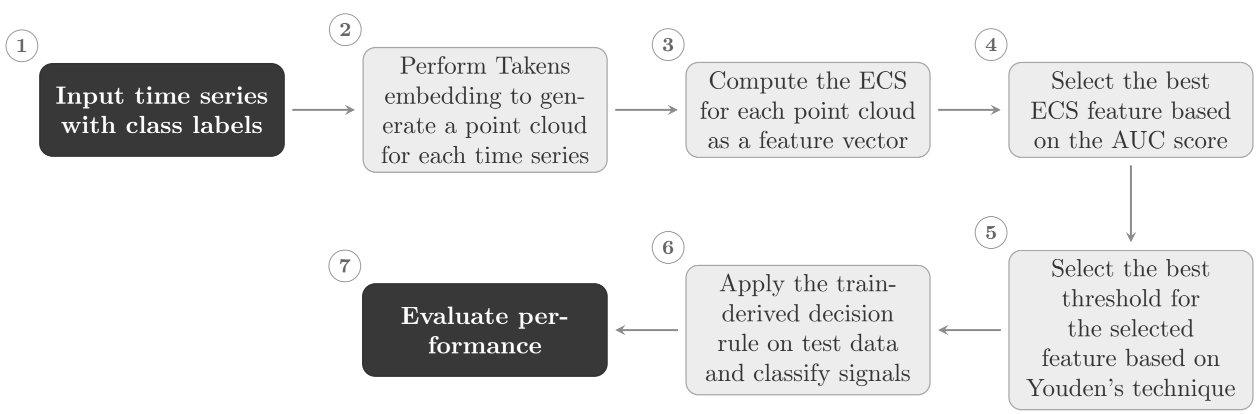


Figure S2: Flowchart of the ECS-based single-feature threshold classification framework.

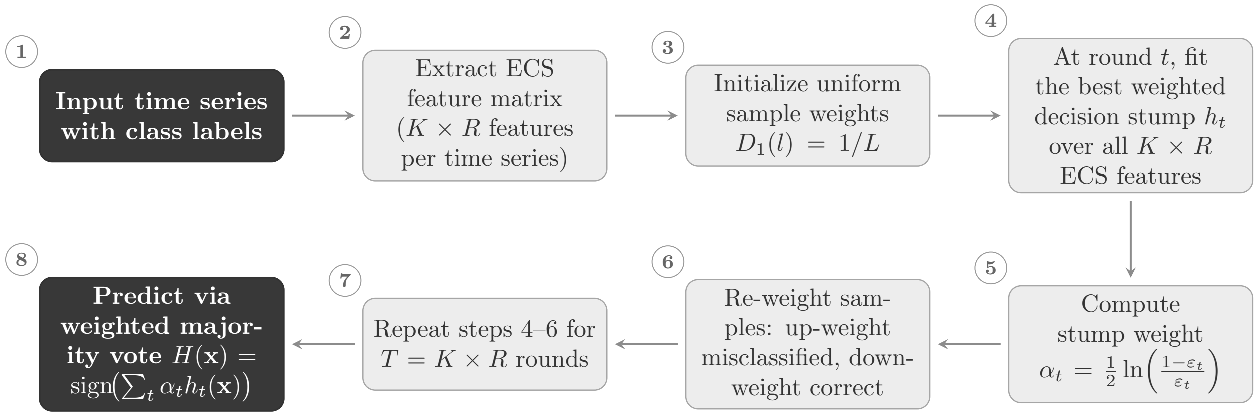


Figure S3: Flowchart of the ECS-based AdaBoost classification framework. The number of weak learners T is set equal to $K \times R$, matching the dimensionality of the ECS feature vector.

$i < j$, each assigned a nonzero multiplicity $\mu^n_{i,j}$, with points on the diagonal included with infinite multiplicity [3].

For $p \geq 1$, the *degree-p Wasserstein distance* between two persistence diagrams $\mathrm{PD}_n(\mathcal{K})$ and $\mathrm{PD}_n(\mathcal{L})$ is:

$$W_p(\mathrm{PD}_n(\mathcal{K}), \mathrm{PD}_n(\mathcal{L})) := \inf_{\varphi} \left(\sum_{x \in \mathrm{PD}_n(\mathcal{K})} d\big(x, \varphi(x)\big)^p \right)^{1/p},$$

where φ is a bijection from $\mathrm{PD}_n(\mathcal{K})$ to $\mathrm{PD}_n(\mathcal{L})$.

Lemma S1 (Stability of Alpha Filtrations in $\mathbb{R}^3$)**.** *Let $\mathcal{D} = \{x_1, \ldots, x_M\}$ be a point cloud in $\mathbb{R}^3$. Then, for an arbitrary small $0 < \epsilon < \frac{1}{2}\min_{1 \leq i < j \leq M} \|x_i - x_j\|_2$, there is $\delta > 0$ such that for any $\mathcal{D}' = \{x'_1, \ldots, x'_M\}$ with $\max_{1 \leq i \leq M} \|x_i - x'_i\|_2 \leq \delta$,*

$$\|\chi(\mathcal{K}(r)) - \chi(\mathcal{K}'(r))\|_1 \leq \frac{1}{4} M^4 \epsilon,$$

where $\mathcal{K}(r)$ and $\mathcal{K}'(r)$ are the corresponding Alpha filtrations on $\mathcal{D}$ and $\mathcal{D}'$, respectively.

Proof. We note that for any non-degenerate simplex (triangle or tetrahedron) T in $\mathbb{R}^3$ and for any $\epsilon > 0$, there is a $\delta_T > 0$ such that after perturbing the vertices by δ_T, the circumradius of T is changed by at most ϵ (this follows from the continuity of the circumradius function of a non-degenerate simplex).

Let $\epsilon > 0$ be less than $\frac{1}{2}\min_{1 \leq i \neq j \leq M} \|x_i - x_j\|_2$. Choose $\delta > 0$ such that for $x'_i \in \mathcal{D}'$ we have $\|x'_i - x_i\| < \min_T \delta_T$, where the minimum is over all non-degenerate simplices formed by points of $\mathcal{D}$. Take the filtering function on a simplex in [4, Theorem 2] to be the radius of the smallest enclosing sphere of the points defining the simplex, with f corresponding to points in $\mathcal{D}$ and g corresponding to points in $\mathcal{D}'$.

Then, by the Cellular Wasserstein theorem of Skraba and Turner, we have for $n = 0, 1, 2$ (noting that the smallest enclosing sphere of a finite set of points in $\mathbb{R}^3$ is realized by two, three, or four of those points):

$$W_1(\mathrm{PD}_0(f), \mathrm{PD}_0(g)) \leq \binom{M}{2}\epsilon,$$

$$W_1(\mathrm{PD}_1(f), \mathrm{PD}_1(g)) \leq \binom{M}{2}\epsilon + \binom{M}{3}\epsilon = \binom{M+1}{3}\epsilon,$$

$$W_1(\mathrm{PD}_2(f), \mathrm{PD}_2(g)) \leq \binom{M}{3}\epsilon + \binom{M}{4}\epsilon = \binom{M+1}{4}\epsilon.$$

Finally, from Theorem 1 of [2]:

$$\begin{aligned}\|\chi(\mathcal{K}(r)) - \chi(\mathcal{K}'(r))\|_1 &\leq 2\sum_{n=0}^{2} W_1(\mathrm{PD}_n(\mathcal{K}), \mathrm{PD}_n(\mathcal{K}')) \\ &= 2\left[2\binom{M}{2} + 2\binom{M}{3} + \binom{M}{4}\right]\epsilon \leq \frac{1}{4}M^4\epsilon.\end{aligned}$$

We sum only over $n = 0, 1, 2$ because the Čech filtration has the same persistent homology as the Alpha filtration. □

Remark. Since this paper uses Takens embedding dimension $m = 3$ for most datasets, we state the stability result here only for $\mathbb{R}^3$. The argument generalizes to $\mathbb{R}^d$ by summing over homological degrees $n = 0, \dots, d-1$ in the Wasserstein bound, yielding a bound that depends on $\binom{M+1}{d+1}$ rather than $\binom{M+1}{4}$.

4 Rössler System: Pairwise L_1 Distance Analysis

Figure S4 shows (a) the pointwise absolute difference $|\mathrm{ECS}_{\mathrm{chaotic}} - \mathrm{ECS}_{\mathrm{periodic}}|$ across the filtration scale–time grid for the Rössler system, and (b) a heatmap of pairwise L_1 distances between ECSs computed from 100 noisy realizations at each noise level from 0% to 10% of the signal amplitude. Intra-regime distances (diagonal blocks) are markedly smaller than inter-regime distances (off-diagonal blocks), confirming that the ECS is robust to moderate noise levels and that the L_1 metric can quantitatively differentiate between the periodic ($c = 2.3$) and chaotic ($c = 7.3$) regimes. The ordering of blocks in panel (b) of Figs. S4, S8, and S10 can be understood with the schematic illustration in Fig. S5.

5 False Nearest Neighbour (FNN) Analysis for Embedding Dimension Selection

To rigorously justify the selection of the embedding dimension m used in the main manuscript, we performed a quantitative False Nearest Neighbour (FNN) analysis across all five datasets. This analysis follows the established methodology of Kennel, Brown, and Abarbanel [5].

For each time series, the FNN ratio was evaluated over embedding dimensions $m \in \{1, \dots, 6\}$ using a time delay of $\tau = 1$. A point is classified as a false nearest neighbour if the relative increase in its nearest-neighbour distance, upon expanding the embedding space to $(m + 1)$th dimensions, exceeds a tolerance threshold of $R_{\mathrm{tol}} = 15$. This evaluation was conducted independently for each class, utilizing signals from the training set only.

The optimal embedding dimension m^* for an individual signal is defined as the smallest integer m at which its FNN ratio drops below a strict 1% threshold. To determine the representative embedding dimension for a given class, we computed the mean of these optimal m^* values across all sampled signals within that class. The final dataset-level embedding dimension reported in the main text is chosen as the integer dimension closest to the overall mean across all classes.

The class-wise mean FNN ratio curves for *ECG5000* and *Epilepsy2* are illustrated in Figure S6. For the *ECG5000* dataset, the mean FNN ratio for both classes falls below the 1% threshold at $m =$

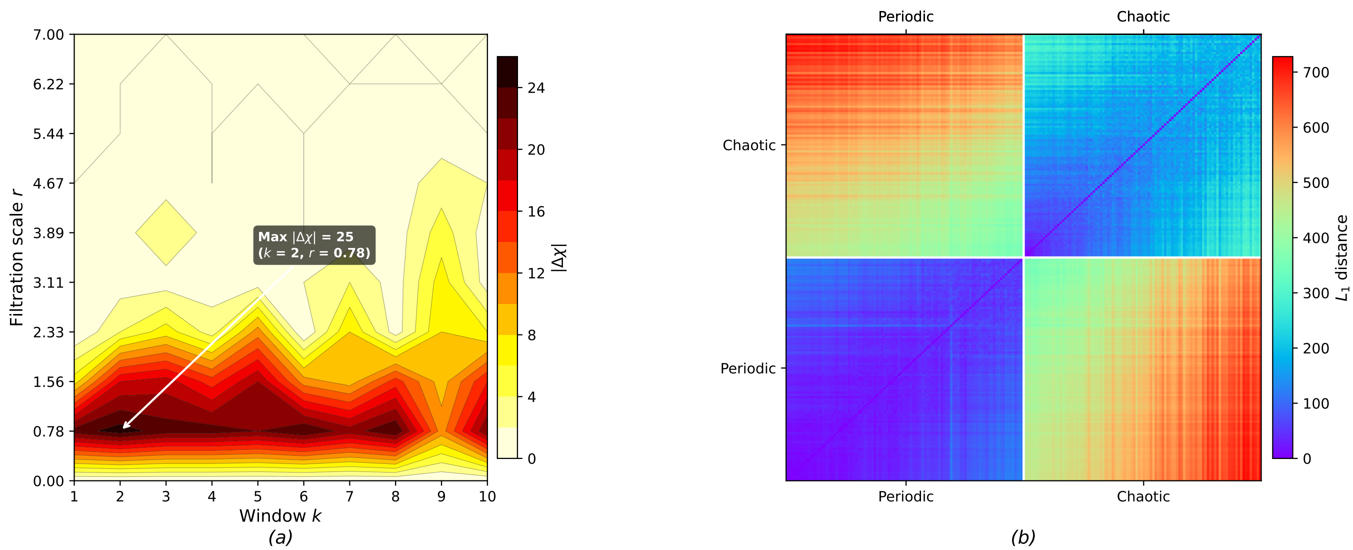


Figure S4: Panel (a) shows the absolute difference $|\text{ECS}_{\text{chaotic}} - \text{ECS}_{\text{periodic}}|$ across the scale–time grid, with the annotated arrow indicating the coordinate of maximum divergence. Panel (b) shows the heatmap of pairwise L_1 distances between ECSs constructed from perturbed time series of the periodic ($c = 2.3$) and chaotic ($c = 7.3$) regimes.

3, with class-specific average optimal dimensions of 3.2 (Class 1) and 3.1 (Class 2). Consequently, $m = 3$ was selected as the uniform embedding dimension. For the *Epilepsy2* dataset, Classes 0 and 1 yielded average optimal dimensions of 3.4 and 3.8, respectively, which justifies the selection of $m = 4$. This same analytical procedure was applied to the three additional ECG datasets (*TwoLeadECG*, *ECG200*, and *ECGFiveDays*), where the FNN analysis consistently supported an embedding dimension of $m = 3$.

6 ECG5000: Grid Size Optimization and ECS Topological Analysis

Grid Size Optimization

Figure S7 shows the mean cross-validated AUC of the best ECS feature as a function of the ECS grid size K for the *ECG5000* dataset. The AUC rises steeply from a 3×3 grid and reaches a peak at the 10×10 configuration, which was therefore adopted for all single-feature experiments.

Topological Comparison of ECS Representations

Figure S8 (a) shows the point-wise absolute difference $|\Delta\chi|$ between the ECSs of two class shown in Fig.3 ((c) and (d)) of the main manuscript. Fig. S8 (b) shows the pairwise L_1 distance heatmap for all training samples. The higher difference in χ is concentrated at window 8 at scale 0.11, corresponding to filtration scale–time coordinate $(r, t) \approx (0.11, 100)$. The mean intra-class L_1 distances for classes 1 and 2 are 5.58 and 4.14, respectively, while the mean inter-class distance is 8.89, confirming topologically meaningful and class-discriminative information encoded in the ECS.

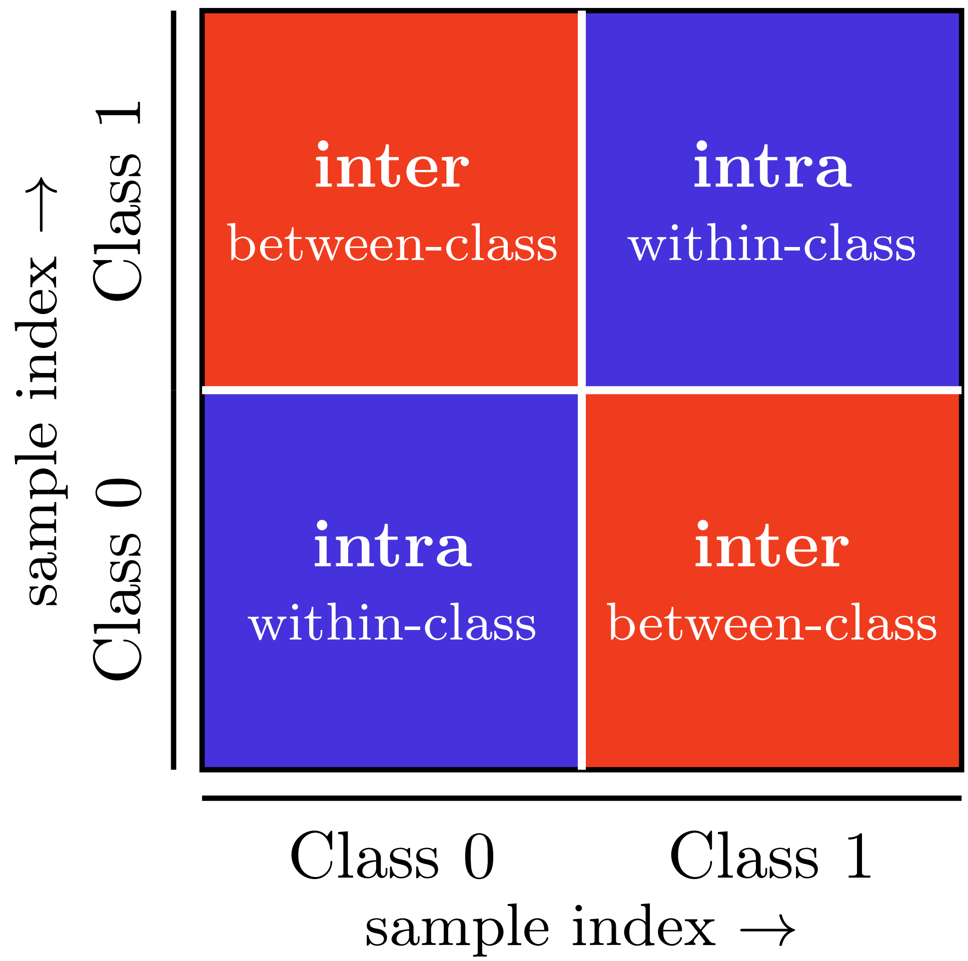
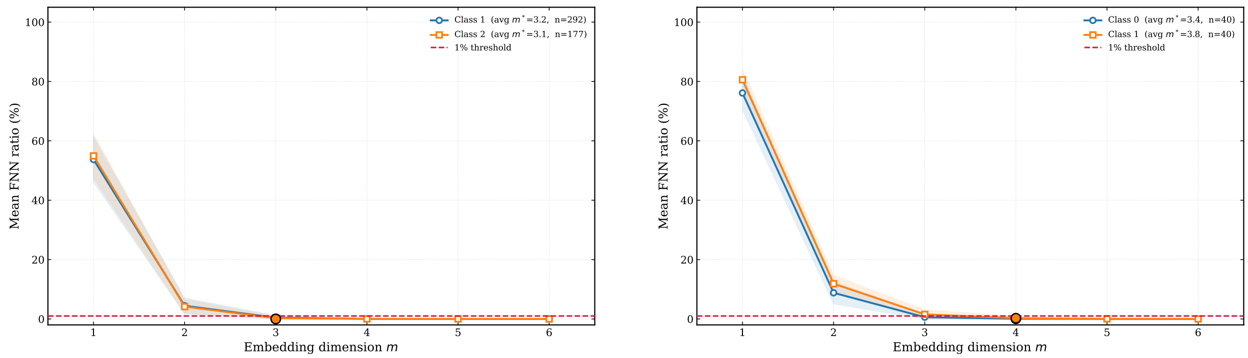


Figure S5: Schematic of the block structure of the class-ordered pairwise distance matrix. Samples are ordered by class along both axes (all Class 0 samples first, then all Class 1 samples). The two *diagonal* blocks contain intra-class (within-regime) distances (blue), while the two *off-diagonal* blocks contain inter-class (between-regime) distances (red). The generic "Class 0" and "Class 1" labels correspond to the specific regime names used in each figure (e.g., Periodic and Chaotic in Fig. S4). This same sample ordering and block interpretation apply to Figs. S4, S8, and S10.



(a) *ECG5000*

(b) *Epilepsy2*

Figure S6: Class-wise False Nearest Neighbour (FNN) ratio curves for *ECG5000* (left) and *Epilepsy2* (right). The curves were computed with a time delay $\tau = 1$ and a distance tolerance $R_{\text{tol}} = 15$. Each line represents the mean FNN ratio per class across the training split, with shaded bands indicating ±1 standard deviation. The dashed horizontal line denotes the 1% selection threshold, and the markers indicate the average optimal embedding dimensions ($m^* = 3$ for *ECG5000* and $m^* = 4$ for *Epilepsy2*).

7 Epilepsy2: ECS Analysis and AdaBoost Contributions

ECS for EEG Signals

Figure S9 shows representative EEG time series and their corresponding ECSs. The seizure ECS (panel (c)) is markedly rougher and more irregular than the smooth, slowly varying non-seizure ECS (panel (d)), reflecting the well-known increase in high-frequency neural oscillations during seizure episodes. This contrast manifests topologically as rapid fluctuations in the connectivity of the Takens point cloud across both filtration scale and time.

Pairwise L_1 Distance Analysis

Figure S10 (a) shows the point-wise absolute difference $|\Delta\chi|$ between the ECSs of the two classes (Fig. S9 (c) and (d)). The differences in χ are spread across multiple windows at scales up to 0.075. Fig. S10 (b) shows the pairwise L_1 distance heatmap between the ECSs constructed from the training samples of the two classes. The pairwise L_1 distances confirm strong inter-class separation (off-diagonal blocks). Intra-class distances within the non-seizure class are very small, indicating high topological regularity, while inter-class distances are substantially larger. The higher intra-class separation (diagonal blocks) for the seizure class may be caused by inherent spatiotemporal differences in the EEG signals of different seizure episodes [6].

AdaBoost α-Contribution Heatmap

Figure S11 shows the α-contribution heatmap for the AdaBoost ensemble trained on *Epilepsy2*. Unlike in the *ECG5000* case, where a single feature dominates, the *Epilepsy2* heatmap reveals multiple features across temporal blocks that contribute substantially to classification. This demonstrates that when discriminatory information is spread across the spatiotemporal range of the ECS rather than concentrated at a single coordinate, the AdaBoost ensemble effectively combines individually weak stumps into a strong classifier.

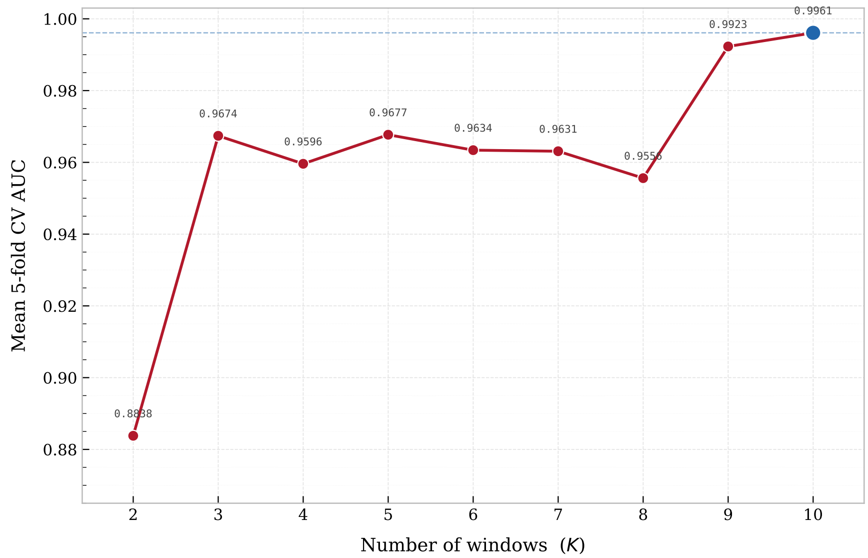


Figure S7: Mean cross-validated AUC of the best ECS feature as a function of the grid size (K) used to construct the ECS for the *ECG5000* dataset.

8 Ablation Studies

To address concerns regarding the attribution of performance gains to specific methodological components, we conducted a comprehensive two-part ablation study across all five datasets. **Part (A)** isolates the architectural contributions of temporal partitioning (K-window construction) and the AdaBoost ensemble extension. **Part (B)** rigorously tests whether the Euler Characteristic Surface (ECS) captures genuine topological information beyond what can be extracted from canonical time-domain statistics or template-matching features within the *same* localized temporal window. All experiments use identical train/test splits and evaluation protocols as the main manuscript.

Component Ablation

We evaluate three configurations of increasing complexity:

- **ECC ($K = 1$)**: A single-scale Euler Characteristic Curve computed over the entire Takens point cloud without temporal partitioning, corresponding to the baseline topological descriptor of Dłotko & Gurnari [7].
- **ECS SF ($K = K^*$)**: The proposed single-feature classifier operating at the optimal scale–time coordinate (r^*, k^*) identified by the ECS pipeline.
- **ECS AdaBoost ($T = K^{*2}$)**: An ensemble of decision stumps trained over the full $K \times R$ grid, aggregating weak topological signals across all coordinates.

Results are summarized in Table S1.

Dataset-specific observations. For *ECG5000*, temporal partitioning improves AUC by +0.034, accuracy by +8.2 pp (0.898→0.980), and F1 by +0.101 (0.872→0.974), with uniform gains across precision, recall, and specificity. The AdaBoost extension contributes only +0.005 AUC, confirming that a single topological coordinate is already sufficient for this dataset.

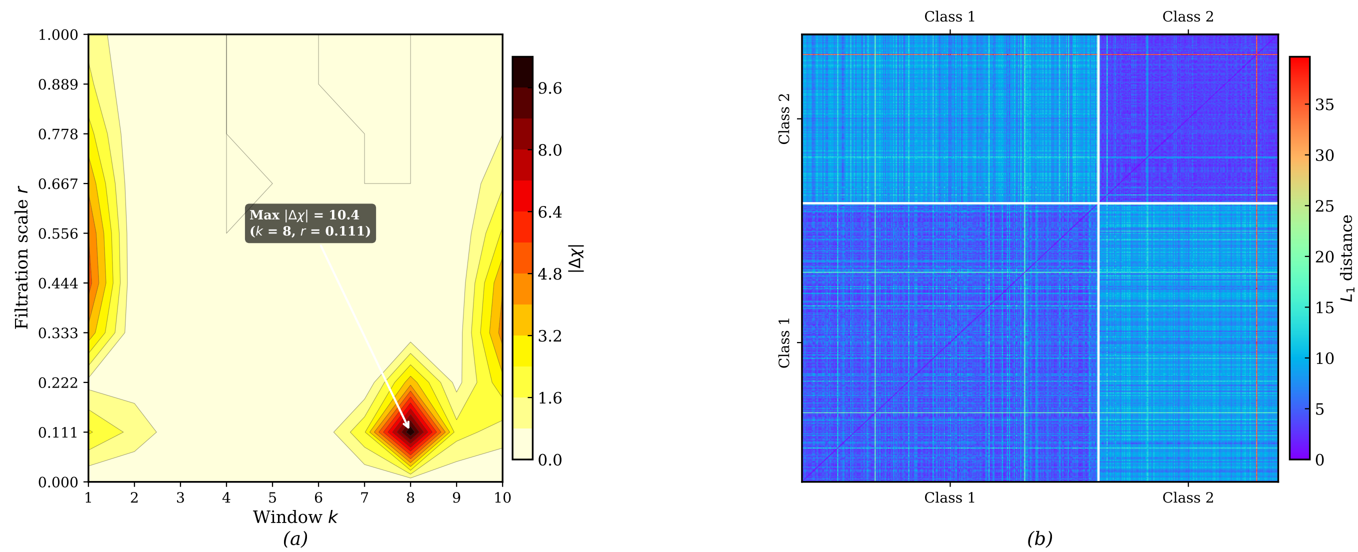


Figure S8: Topological comparison of ECS representations from the *ECG5000* training set. Panel (a) shows the pointwise absolute difference $|\Delta\chi|$ between the ECSs of each class over the scale–window grid. Panel (b) shows the pairwise L_1 distances between the ECSs of all training samples, visualized as a heatmap.

Dataset	Configuration	AUC	Accuracy	F1 Score	Precision	Recall	Specificity
ECG5000	ECC ($K = 1$)	0.9593	0.8980	0.8723	0.8262	0.9239	0.8824
	ECS SF ($K^* = 10$)	0.9937	0.9803	0.9736	0.9852	0.9623	0.9912
	ECS AdaBoost ($T = 36$)	**0.9988**	**0.9855**	**0.9807**	**0.9848**	**0.9767**	**0.9909**
Epilepsy2	ECC ($K = 1$)	0.9578	0.8553	0.9028	**0.9792**	0.8374	**0.9279**
	ECS SF ($K^* = 9$)	0.8347	0.8761	0.9212	0.9398	0.9034	0.7655
	ECS AdaBoost ($T = 81$)	**0.9628**	**0.9256**	**0.9531**	0.9631	**0.9433**	0.8535
TwoLeadECG	ECC ($K = 1$)	0.8484	0.8490	0.8317	0.9403	0.7456	0.9525
	ECS SF ($K^* = 9$)	0.9742	0.9263	0.9296	0.8894	**0.9737**	0.8787
	ECS AdaBoost ($T = 81$)	**0.9780**	**0.9412**	**0.9391**	**0.9736**	0.9070	**0.9754**
ECG200	ECC ($K = 1$)	0.7014	0.7400	0.7969	0.7969	0.7969	0.6389
	ECS SF ($K^* = 9$)	0.8516	0.7800	0.8000	**0.9565**	0.6875	**0.9444**
	ECS AdaBoost ($T = 81$)	**0.8811**	**0.8000**	**0.8507**	0.8143	**0.8906**	0.6389
ECGFiveDays	ECC ($K = 1$)	0.6093	0.5668	0.5906	0.5628	0.6212	0.5117
	ECS SF ($K^* = 10$)	0.7949	0.7851	**0.8038**	0.7431	**0.8753**	0.6939
	ECS AdaBoost ($T = 100$)	**0.8494**	**0.7933**	0.7959	**0.7904**	0.8014	**0.7850**

Table S1: **Part (A) ablation — component attribution.** Three configurations in order of increasing complexity across all five datasets. ECC ($K = 1$) serves as the single-scale topological baseline [7]. Bold denotes the best value per metric per dataset.

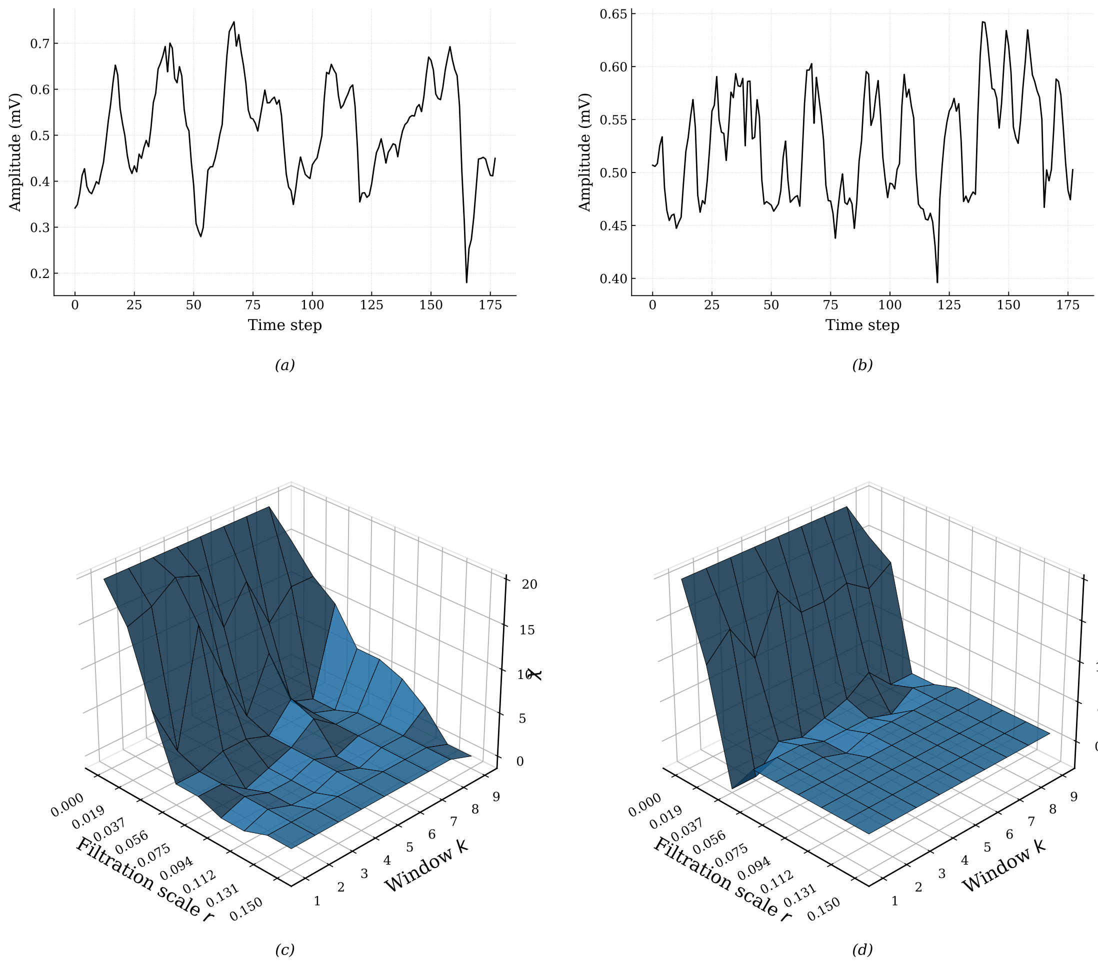


Figure S9: Panels (a) and (b) show representative EEG time series for the seizure and non-seizure classes, respectively. Panels (c) and (d) show the corresponding ECSs.

Epilepsy2 reveals a complementary regime. The global ECC ($K = 1$) actually surpasses the single ECS feature in AUC (0.958 vs 0.835), precision (0.979 vs 0.940), and specificity (0.928 vs 0.766). This occurs because ictal events modify the *global* attractor geometry of the EEG trajectory; temporal partitioning fragments rather than concentrates the discriminant, leaving the isolated coordinate unable to capture the full topological shift. Here, **AdaBoost is architecturally essential**: it elevates AUC from 0.835 to 0.963 (+0.128), accuracy by +4.9 pp ($0.876 \rightarrow 0.926$), F1 by +0.032, and recall from 0.903 to 0.943 (+0.04 pp), successfully aggregating distributed evidence across all 81 scale–time coordinates to recover and exceed the global baseline.

For *TwoLeadECG*, the K-window yields a +0.126 AUC improvement, driven primarily by a +22.8 pp recall gain ($0.746 \rightarrow 0.974$). This indicates that the global EC curve systematically misses true positives that the localized window construction successfully isolates. AdaBoost adds a marginal +0.004 AUC, consistent with a highly concentrated discriminant.

In *ECG200*, the +0.150 AUC gain from partitioning is accompanied by a precision–recall tradeoff. ECS SF achieves the highest precision (0.957) and specificity (0.944) among all configurations, reflecting a conservative decision boundary. AdaBoost corrects this by raising recall from 0.688 to 0.891 with only a modest specificity reduction.

Finally, *ECGFiveDays* exhibits the largest localization benefit: AUC increases by +0.186, ac-

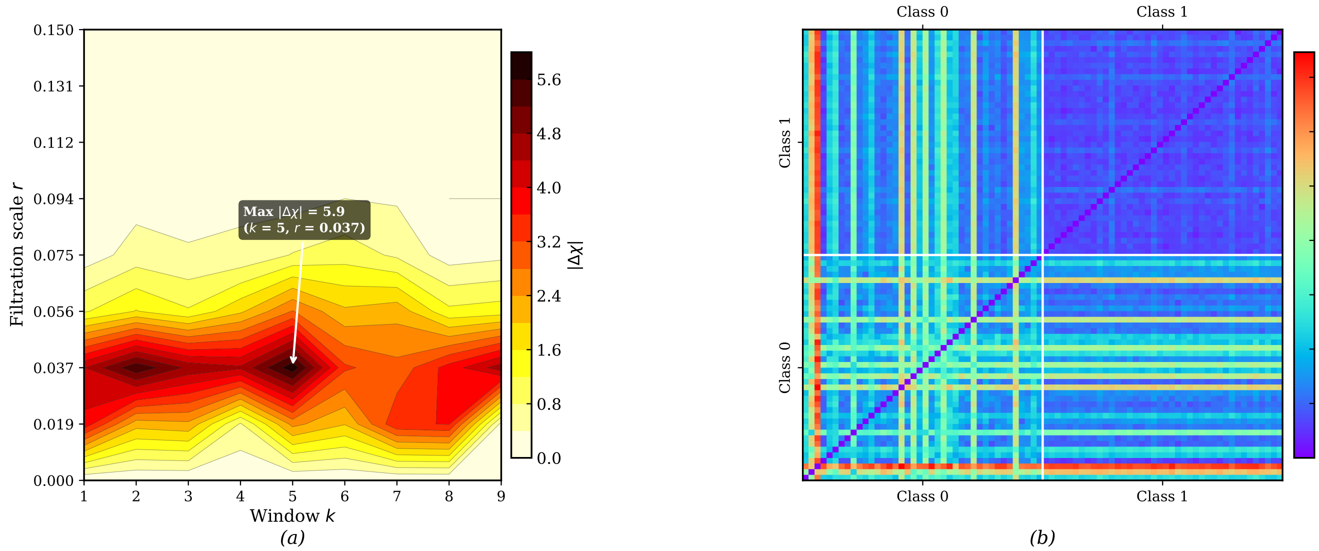


Figure S10: Topological comparison of ECS representations from the *Epilepsy2* training set. Panel (a) shows the point-wise absolute difference $|\Delta\chi|$ between the ECS of each class over the scale–window grid. Panel (b) shows the pairwise L_1 distance between the ECSs of all training samples, displayed as a heatmap.

curacy by +21.8 pp ($0.567 \to 0.785$), and F1 by +21.3 pp ($0.591 \to 0.804$). The near-chance ECC baseline confirms that no class-discriminative structure exists at the global topological level.

These results establish a clear design principle. **The K-window construction is the dominant architectural driver when the discriminant is temporally localized**, delivering AUC gains of +0.034 to +0.186 by isolating the sub-region where topological signal concentrates. The AdaBoost extension is indispensable when the discriminant is distributed across the scale–time plane, yielding its largest improvement precisely where a single coordinate is least informative (*Epilepsy2*: +0.128 AUC) and its smallest where a single coordinate already suffices (*ECG5000*, *TwoLeadECG*: +0.005AUC, +0.004 AUC). This complementary behavior validates the two-tier framework as a principled adaptation to the two qualitatively distinct regimes of biomedical time-series discriminability: localized morphological markers and distributed dynamical transitions.

Topological Content vs. Localized Time-Domain/Shapelet Features

To determine whether the ECS merely acts as a surrogate for amplitude-based descriptors, we compare $\chi(r^*, k^*)$ (B6) against five features extracted from the *exact same time steps* identified by the ECS pipeline:

- B1–B3: Canonical interval-based statistics (mean, standard deviation, amplitude range) as formalized in the Time Series Forest [8] and TSC taxonomy [9].
- B4: AdaBoost ensemble over B1–B3.
- B5: Template-matching distance (ℓ^2 norm to the class-conditional mean sub-sequence within window k^*).
- B6: ECS feature $\chi(r^*, k^*)$.

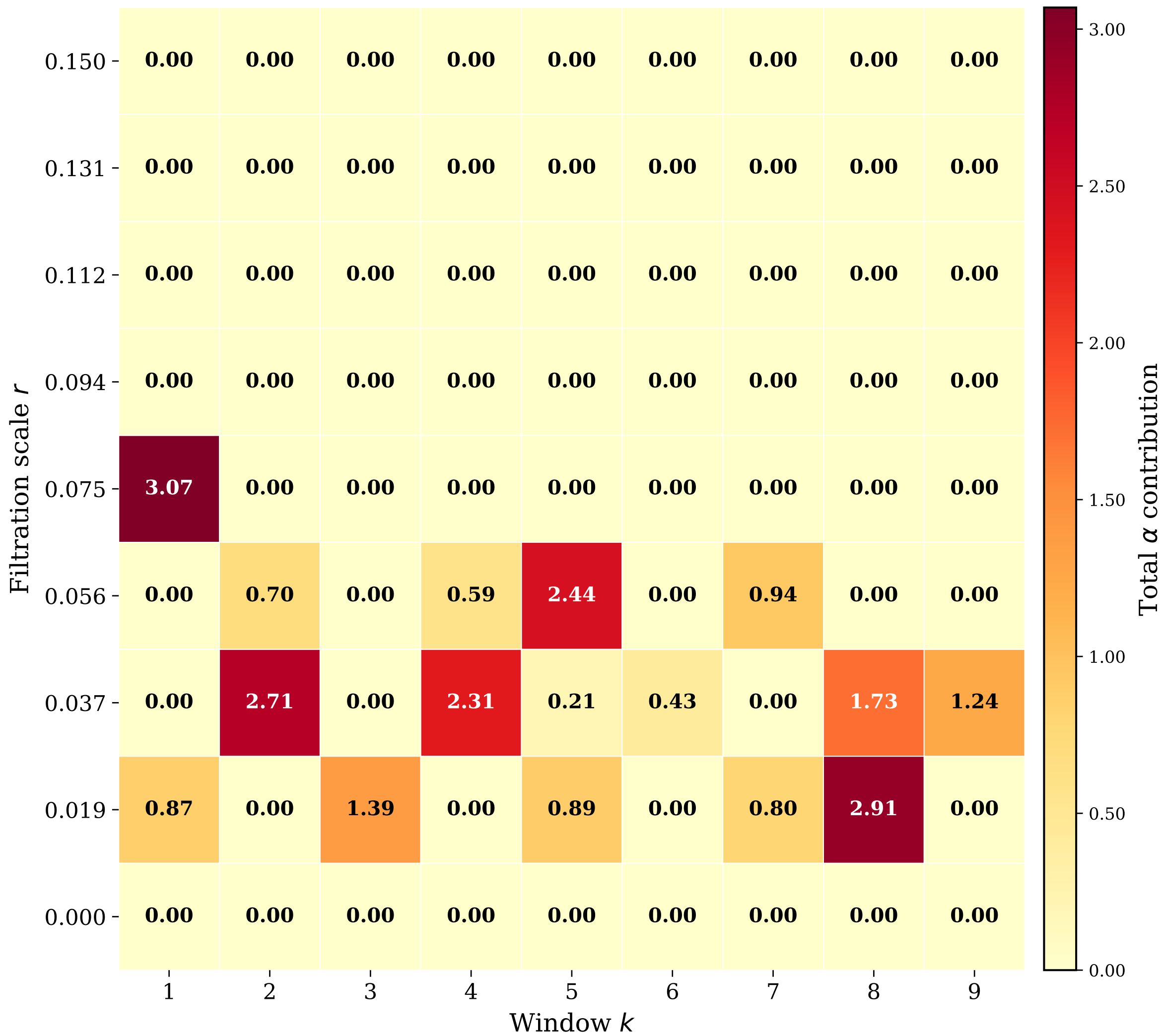


Figure S11: Heatmap of the total α contribution of each ECS feature in the AdaBoost ensemble trained on the *Epilepsy2* dataset (9×9 grid, $T = 81$ stumps).

- B7: ECS AdaBoost.

B5 provides a direct test of whether average waveform shape in the identified window is discriminative, but it is intentionally conservative compared to the full Shapelet Transform Classifier (already evaluated in the main baselines). Crucially, because configurations B1–B5 all receive the ground-truth window k^* discovered by ECS during training, any residual advantage of B6 represents a *conservative lower bound* on the true topological added value. Results are presented in Table S2.

Dataset-specific observations. For *ECG5000*, standard deviation (B2, AUC ≈ 0.990) and range (B3, AUC ≈ 0.990) are already competitive with the single ECS feature (B6, AUC 0.9937), trailing by only 0.003–0.004 AUC. This indicates that the strong performance is substantially associated with a localized ST–T/T-wave morphology difference. The single ECS feature offers a modest but consistent incremental contribution. It edges ahead of every individual amplitude statistic across the six metrics and surpasses the template/shapelet distance (B5, AUC 0.974), while Theorem 1 additionally guarantees a stability under perturbation that variance and range lack. We therefore interpret *ECG5000* as a case where ECS captures a topological manifestation of a strong localized waveform difference, not as evidence of a decisive, purely topological advantage.

For *Epilepsy2*, mean amplitude (B1) is near-random, while amplitude variability (B2) and range (B3) exceed the single ECS feature (B6) on AUC and F1. The single coordinate nonetheless attains the highest precision (0.940) and specificity (0.766) among these baselines. The stronger performance of ECS AdaBoost indicates that the discriminant on this dataset is distributed across the filtration-scale–time grid rather than captured by any single coordinate.

TwoLeadECG is the most decisive case. Mean amplitude (B1) is near-random, ruling out amplitude-level differences. Standard deviation (B2) lags ECS-SF by 0.137 AUC, range (B3) by 0.169, and multi-statistic AdaBoost (B4) by 0.156 AUC, while the template-matching distance (B5) also fails. The discriminant resides exclusively in the topological structure of the Takens-embedded trajectory at filtration scale $r^* \approx 0.144$, indicating that ECS extracts genuine topological information rather than acting as a surrogate amplitude-variance detector.

For *ECG200*, the single ECS feature (B6) does not improve on raw amplitude variability and is in fact marginally behind on AUC (0.852 vs. B3's 0.887). The discriminant is not cleanly topologically localized, and a single ECS coordinate is correspondingly modest. It is interesting to note that the amplitude statistics themselves only reach the mid-0.88s. Therefore, the dataset offers little discriminative signal of any kind in this window, topological or otherwise. For *ECGFiveDays*, the picture is similar to that of *ECG200*: the single ECS feature (B6, AUC 0.795) trails the raw amplitude features. The modest results for *ECG200* and *ECGFiveDays* by the single ECS feature classifier are improved only partially by the AdaBoost ensemble.

Overall, the ECS framework extracts genuine topological structure complementary to traditional time-domain descriptors. Because B1–B5 receive the ground-truth window k^* discovered by ECS, any residual advantage of the single ECS feature is a conservative lower bound on its topological added value.

9 Comparison with Representative Baselines

To contextualise the ECS framework against established time-series classification (TSC) benchmarks, we evaluated several representative methods spanning the major TSC paradigms. All methods were implemented under identical experimental conditions (standard UCR train/test splits [10], unmodified data loading, and no additional preprocessing). The baselines include:

1. **1-Nearest Neighbour with Dynamic Time Warping (1-NN DTW)** [11, 12]: The standard TSC benchmark for over two decades [9]. It is fully interpretable but cannot capture global topological or phase-space structure.

2. **Shapelet Transform Classifier (STC)** [13, 14]: Learns discriminative subsequences of the raw signal. STC is interpretable in the amplitude domain but operates exclusively on raw amplitude and cannot access phase-space topology.

3. **RandOm Convolutional KErnel Transform (ROCKET)** [15]: Applies 10,000 randomly generated convolutional kernels and trains a ridge regression classifier. Achieves near-state-of-the-art accuracy but relies on a massive random feature space with no individual feature interpretable in terms of signal dynamics.

4. **MiniROCKET** [16]: A nearly deterministic reformulation of ROCKET, up to 75× faster while retaining essentially identical accuracy. Shares ROCKET's interpretability limitations.

9.1 ECG5000 Dataset

Literature-based comparison with the PH-based baseline under closely matched protocol settings. To contextualize our results against the persistent-homology pipeline of Ichinomiya [17] under closely matched settings, we present a step-by-step architectural breakdown in Table S3. Both methods use the same UCR archive, identical binary class pair (classes 1 and 2), the same pre-normalized time series of length $L = 140$, and the same Takens embedding parameters ($m = 3$, $\tau = 1$). The key difference is the topological computation: Ichinomiya converts the embedded point cloud into a binary recurrence plot and computes persistent homology ($O(n^{\omega})$), followed by persistence-image vectorization and NMF reduction; our method applies an Alpha-complex filtration directly to the point cloud and computes the Euler Characteristic Surface ($O(n + R{\cdot}T)$), requiring no vectorization or dimensionality reduction. Both pipelines conclude with a single-feature threshold classifier. The comparison therefore provides contextual evidence under closely matched protocol settings, while remaining literature-based rather than a fully reimplemented head-to-head benchmark.

9.2 Statistical Significance Testing

To assess whether the performance differences reported in Table 4 of the main manuscript are statistically significant, we conducted paired significance tests for every ECS-vs-baseline comparison. Because all classifiers are evaluated on the identical held-out UCR test split of each dataset, the comparisons are naturally paired on the same test instances. We assess differences in AUC with the DeLong test for two correlated ROC curves [19] (using the fast algorithm of Sun and Xu [20]), and differences in accuracy with McNemar's test [21], which Dietterich [22] recommends for classifiers trained once on a fixed test split. Because McNemar's test is symmetric—a significant p-value indicates that two classifiers differ, not which is superior—we also report the sign of the accuracy difference so that the direction of every result is explicit. Within each dataset, p-values across the four baselines are Holm–Bonferroni corrected [23]. Results for the ECS AdaBoost classifier are reported in Table S4.

The tests support a consistent and honestly-stated conclusion. On raw predictive metrics, the shapelet- and convolution-based baselines generally outperform ECS: their AUC is significantly higher than ECS AdaBoost on *TwoLeadECG*, *ECGFiveDays*, *ECG5000*, and *Epilepsy2* (DeLong $p < 0.001$ in each case), and they are significantly more accurate on most datasets. On *ECG5000*, where all methods operate near ceiling, the AUC gap is very small ($|\Delta\text{AUC}| \leq 0.001$) yet statistically detectable given the large test set ($n = 4{,}217$), and the baselines hold a small but significant accuracy advantage (e.g. ROCKET $\Delta\text{Acc} = -0.009$, McNemar $p < 10^{-5}$). On *ECGFiveDays* and *Epilepsy2*, the AUC advantage of the baselines is larger. Relative to the distance-based 1-NN DTW baseline, ECS is more accurate on *TwoLeadECG*, *ECG200*, and *ECGFiveDays*, though this is not uniform across datasets and is not the basis of our claims.

We emphasize that ECS does not lead the benchmark on raw predictive metrics. Its contribution, as detailed throughout the main manuscript, is orthogonal to ranking performance—every ECS prediction is traceable to a specific coordinate (r^*, k^*) in the Euler Characteristic Surface, providing a formally stable (Theorem 1), filtration scale–time-localized topological explanation that none of the baselines offer. On *ECG5000*, for example, this coordinate maps automatically to the ST–T segment (time steps 98–114) with no clinical prior. The statistical tests quantify precisely the modest cost in AUC and accuracy at which this interpretability is obtained.

Dataset	Configuration	AUC	Accuracy	F1	Precision	Recall	Specificity
Window $k^*=8$, filtration scale $r^*\approx 0.111$							
ECG5000	B1: Mean	0.9033	0.8577	0.8338	0.7450	0.9465	0.8040
	B2: Std	0.9900	0.9708	0.9612	0.9634	0.9591	0.9779
	B3: Range	0.9904	0.9749	0.9664	0.9738	0.9591	0.9844
	B4: Multi+Ada	0.9942	0.9760	0.9680	0.9751	0.9610	0.9852
	B5: Template dist.	0.9738	0.9322	0.9161	0.8582	**0.9824**	0.9018
	B6: ECS χ	0.9937	0.9803	0.9736	0.9852	0.9623	0.9912
	B7: ECS AdaBoost	**0.9988**	**0.9855**	**0.9807**	**0.9848**	**0.9767**	**0.9909**
Window $k^*=9$, filtration scale $r^*\approx 0.037$							
Epilepsy2	B1: Mean	0.4781	0.7567	0.8491	0.8446	0.8537	0.3633
	B2: Std	0.8738	0.9147	0.9485	0.9201	**0.9787**	0.6553
	B3: Range	0.8957	0.9067	0.9422	0.9353	0.9492	0.7341
	B4: Multi+Ada	0.8801	0.9076	0.9429	0.9345	0.9515	0.7296
	B5: Template dist.	0.6335	0.5891	0.6950	0.8589	0.5836	0.6115
	B6: ECS χ	0.8347	0.8761	0.9212	0.9398	0.9034	0.7655
	B7: ECS AdaBoost	**0.9628**	**0.9256**	**0.9531**	**0.9631**	0.9433	**0.8535**
Window $k^*=5$, filtration scale $r^*\approx 0.144$							
TwoLeadECG	B1: Mean	0.5748	0.5803	0.3856	0.7212	0.2632	0.8981
	B2: Std	0.8368	0.7629	0.7536	0.7852	0.7246	0.8014
	B3: Range	0.8049	0.7270	0.6868	0.8061	0.5982	0.8559
	B4: Multi+Ada	0.8180	0.7278	0.6888	0.8052	0.6018	0.8541
	B5: Template dist.	0.6733	0.6067	0.3760	0.9122	0.2368	**0.9772**
	B6: ECS χ	0.9742	0.9263	0.9296	0.8894	**0.9737**	0.8787
	B7: ECS AdaBoost	**0.9780**	**0.9412**	**0.9391**	**0.9736**	0.9070	0.9754
Window $k^*=4$, filtration scale $r^*\approx 0.206$							
ECG200	B1: Mean	0.6198	0.7400	0.8060	0.7714	0.8438	0.5556
	B2: Std	0.8832	0.7700	0.8000	0.9020	0.7188	0.8611
	B3: Range	**0.8867**	0.7700	0.7965	0.9184	0.7031	0.8889
	B4: Multi+Ada	0.8843	**0.8200**	**0.8657**	0.8286	**0.9062**	0.6667
	B5: Template dist.	0.8655	0.7700	0.8130	0.8475	0.7812	0.7500
	B6: ECS χ	0.8516	0.7800	0.8000	**0.9565**	0.6875	**0.9444**
	B7: ECS AdaBoost	0.8811	0.8000	0.8507	0.8143	0.8906	0.6389
Window $k^*=7$, filtration scale $r^*\approx 0.389$							
ECGFiveDays	B1: Mean	0.8527	0.7456	0.7812	0.6884	0.9030	0.5864
	B2: Std	**0.8887**	0.7236	0.6394	**0.9295**	0.4873	**0.9626**
	B3: Range	0.8799	0.7805	0.7629	0.8352	0.7021	0.8598
	B4: Multi+Ada	0.8629	0.7805	0.7664	0.8245	0.7159	0.8458
	B5: Template dist.	0.8615	0.7666	0.7759	0.7500	0.8037	0.7290
	B6: ECS χ	0.7949	0.7851	**0.8038**	0.7431	**0.8753**	0.6939
	B7: ECS AdaBoost	0.8494	**0.7933**	0.7959	0.7904	0.8014	0.7850

Table S2: **ECS topological feature vs raw amplitude features from the same window k^*.** B1–B5 receive the ground-truth window k^*; any residual gap in favour of B6 or B7 is a conservative lower bound on topological added value. B6 = ECS feature $\chi(r^*, k^*)$; B7 = ECS AdaBoost. Bold: best value per metric per dataset.

Aspect	Ichinomiya (2025) [17]	This work
Data & setup	*ECG5000* (UCR/UEA archive [18]); classes 1 & 2; 469 train / 4,217 test (binary subset); $L = 140$	*ECG5000* (UCR/UEA archive [18]); classes 1 & 2; 469 train / 4,217 test (binary subset); $L = 140$
Embedding	UCR pre-normalized input; Takens, m=3, τ=1	UCR pre-normalized input; Takens, m=3, τ=1 (FNN criterion [5])
Intermediate step	Recurrence plot (binary $N \times N$ matrix, distance threshold ε)	*None* — filtration applied directly to the point cloud
Topological computation	Persistent homology on recurrence plot ; $O(n^{\omega})$, $2 \le \omega < 2.373$	Alpha-complex filtration; ECS; $O(n + R \cdot T)$
Feature representation	Persistence diagram $\xrightarrow{\text{kernel}}$ persistence image $\xrightarrow{\text{NMF}}$ k components	ECS integer matrix $K \times R$; no vectorization or dimension reduction
Feature selection	Fixed: NMF component 1; **6 pipeline steps**	AUC over $K^* \times R$=100 cells, training set only; **3 pipeline steps**
Classifier	Single-feature threshold; threshold rule not reported	Single-feature threshold; Youden index (training ROC)
Performance	Accuracy 62%; AUC 0.90	Accuracy **98.0%**; AUC **0.994**
Feature interpretability	NMF 1: abstract mixture of persistence-image pixels; no direct time or filtration scale correspondence	$\chi(r^* \approx 0.111, k^*=8)$: maps to time steps 98–114 (ST–T segment); identified from training data alone

Table S3: Step-by-step architectural comparison of the PH-based pipeline of Ichinomiya [17] and the proposed ECS pipeline on *ECG5000* (binary, classes 1 & 2, standard UCR split). Both use an identical single-feature threshold classifier; the comparison provides contextual evidence under closely matched protocol settings, while remaining literature-based rather than a fully reimplemented head-to-head benchmark.

Dataset	**Baseline**	**ΔAUC**	**DeLong** p	**ΔAcc**	**McNemar** p	**More acc.**
ECG5000	1-NN DTW	$+0.0099$	n/a	-0.0052	0.016	baseline
	Shapelet Transform	-0.0010	8.53×10^{-4}	-0.0071	1.50×10^{-3}	baseline
	ROCKET	-0.0011	4.59×10^{-4}	-0.0088	8.33×10^{-6}	baseline
	MiniROCKET	-0.0011	8.53×10^{-4}	-0.0092	4.57×10^{-6}	baseline
Epilepsy2	1-NN DTW	$+0.0874$	n/a	-0.0176	4.16×10^{-11}	baseline
	Shapelet Transform	-0.0254	$< 10^{-16}$	-0.0034	0.244	baseline
	ROCKET	-0.0257	$< 10^{-16}$	-0.0362	1.13×10^{-42}	baseline
	MiniROCKET	-0.0266	$< 10^{-16}$	-0.0370	6.56×10^{-49}	baseline
TwoLeadECG	1-NN DTW	$+0.0736$	n/a	$+0.0369$	1.66×10^{-3}	ECS
	Shapelet Transform	-0.0220	6.13×10^{-7}	-0.0588	2.97×10^{-15}	baseline
	ROCKET	-0.0220	6.13×10^{-7}	-0.0579	9.63×10^{-15}	baseline
	MiniROCKET	-0.0219	6.13×10^{-7}	-0.0571	2.62×10^{-14}	baseline
ECG200	1-NN DTW	$+0.1337$	n/a	$+0.0300$	0.767	ECS
	Shapelet Transform	-0.0317	0.323	-0.0500	0.767	baseline
	ROCKET	-0.0855	9.91×10^{-3}	-0.1200	0.0167	baseline
	MiniROCKET	-0.0816	0.0136	-0.1100	0.0222	baseline
ECGFiveDays	1-NN DTW	$+0.0809$	n/a	$+0.0256$	0.203	ECS
	Shapelet Transform	-0.1506	$< 10^{-16}$	-0.2067	1.44×10^{-39}	baseline
	ROCKET	-0.1506	$< 10^{-16}$	-0.2067	1.44×10^{-39}	baseline
	MiniROCKET	-0.1506	$< 10^{-16}$	-0.2067	1.44×10^{-39}	baseline

Table S4: Statistical comparison of ECS AdaBoost against each benchmark baseline on held-out UCR test splits. ΔAUC and ΔAcc are (ECS − baseline); negative values indicate the baseline is higher. DeLong (AUC) and McNemar (accuracy) p-values are Holm–Bonferroni corrected across the four baselines per dataset. As McNemar's test is symmetric, the "More acc." column names the more accurate method. For 1-NN DTW, which is hard-scoring (two-valued output), the AUC point estimate and ΔAUC are reported for reference, but the DeLong significance test is not applicable ("n/a").

References

[1] Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & computational geometry*, 28(4):511–533, 2002.

[2] Anamika Roy, Atish J Mitra, and Tapati Dutta. Euler characteristic surfaces: A stable multiscale topological summary of time series data. *La Matematica*, page 735–763, 2025.

[3] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17, 2017.

[4] Primoz Skraba and Katharine Turner. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*, 2020.

[5] Matthew B Kennel, Reggie Brown, and Henry DI Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6):3403, 1992.

[6] Mariella Panagiotopoulou, Christoforos A. Papasavvas, Gabrielle M. Schroeder, Rhys H. Thomas, Peter N. Taylor, and Yujiang Wang. Fluctuations in eeg band power at subject-specific timescales over minutes to days explain changes in seizure evolutions. *Human Brain Mapping*, 43(8):2460–2477, 2022.

[7] Paweł Dłotko and Davide Gurnari. Euler characteristic curves and profiles: a stable shape invariant for big data problems. *GigaScience*, 12:giad094, 2023.

[8] Houtao Deng, George Runger, Eugene Tuv, and Martyanov Vladimir. A time series forest for classification and feature extraction. *Information Sciences*, 239:142–153, 2013.

[9] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data mining and knowledge discovery*, 31(3):606–660, 2017.

[10] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

[11] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[12] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

[13] Jason Lines, Luke M. Davis, Jon Hills, and Anthony Bagnall. A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 289–297, 2012.

[14] Jon Hills, Jason Lines, Edgaras Baranauskas, James Mapp, and Anthony Bagnall. Classification of time series by shapelet transformation. *Data mining and knowledge discovery*, 28(4):851–881, 2014.

[15] Angus Dempster, François Petitjean, and Geoffrey I. Webb. ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, September 2020.

[16] Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb. Minirocket: A very fast (almost) deterministic transform for time series classification. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 248–257, 2021.

[17] Takashi Ichinomiya. Machine learning of time series data using persistent homology. *Scientific Reports*, 15(1):20508, 2025.

[18] Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*, 2018.

[19] Elizabeth R DeLong, David M DeLong, and Daniel L Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837–845, 1988.

[20] Xu Sun and Weichao Xu. Fast implementation of delong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Processing Letters*, 21(11):1389–1393, 2014.

[21] Quinn McNemar. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, 1947.

[22] Thomas G Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923, 1998.

[23] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2):65–70, 1979.