Scilight

Journal of Machine Learning and Information Security
https://www.sciltp.com/journals/jmlis

Article

Check for updates

DP-SoftShape: Adaptive Differential Privacy via Attention-Guided Sparsification for Time-Series Classification

Baobing Zhang [1,*], Wanxin Sui [2] and Maozhen Li [2,*]

[1] Robotics Research Group, School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK
[2] Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, UK
* Correspondence: b.zhang6@herts.ac.uk (B.Z.); Maozhen.Li@brunel.ac.uk (M.L.)

How To Cite: Zhang, B.; Sui, W.; Li, M.; et al. DP-SoftShape: Adaptive Differential Privacy via Attention-Guided Sparsification for Time-Series Classification. *Journal of Machine Learning and Information Security* **2026**, *2*(2), 9. https://doi.org/10.53941/jmlis.2026.100009

Received: 25 January 2026
Revised: 13 May 2026
Accepted: 19 May 2026
Published: 27 May 2026

Abstract: Local Differential Privacy (LDP) is the de facto protocol for releasing time-series data under formal privacy guarantees. Existing LDP schemes typically add Laplace noise of equal scale at every time step, which tends to drown out the short discriminative patches a classifier depends on. DP-SoftShape spends ϵ unevenly across patches. An attention head scores each shape patch by its contribution to the class label and the Laplace scale at that patch is set in inverse proportion to the score, so noisier perturbation falls on flatter regions while the patches that actually carry the class signal stay close to the original embedding. A uniform-budget version of the same architecture loses most of this gain (Section 4). A Mixture-of-Experts refinement layer downstream recovers features that the noise still disturbs. On 20 UCR datasets at four privacy budgets, DP-SoftShape attains a mean accuracy of 0.722, against 0.565 for ROCKET and 0.553 for Arsenal under the same input-LDP setting.

Keywords: differential privacy; time series classification; adaptive noise allocation; mixture-of-experts

1. Introduction

Smart meters now log household power usage at sub-second resolution [1], and wrist-worn devices stream cardiac rhythms throughout the day [2]; similar sampling rates appear in payment scoring and other client-side telemetry. The downstream task in all of these cases is a Time Series Classification (TSC) problem [3]. TSC accuracy on UCR [4] has improved sharply with deep models such as InceptionTime [5] and ROCKET [6], but the fine-grained traces that drive that accuracy also encode the person behind them, gait, sleep pattern, household routine. The moment those signals leave the device, the trace stops being a feature vector and starts being identifying data.

Concerns about that kind of leakage are not abstract. A handful of timestamped readings have been shown to be enough to re-identify an individual in mobility and consumption datasets [7–9], and recent regulation, GDPR [10,11] in the EU and HIPAA [12] in the US, has made it hard, in many cases unlawful, to ship raw user-level traces to a central server [13]. Local Differential Privacy (LDP) [14] is the response that has stuck in this setting: each client perturbs its own data before anything leaves the device, removing the need to trust a central aggregator [15,16]. Compared to alternatives, LDP has two practical advantages we care about. k-anonymity is brittle once auxiliary information is available, and homomorphic encryption is too expensive at the sampling rates we see in temporal sensors; LDP, by contrast, gives a quantifiable bound on per-release information leakage, and the bound composes cleanly across releases via the sequential composition theorem [13].

What hurts in practice is what LDP does to classification accuracy. Almost every existing temporal LDP scheme adds Laplace noise of the same scale to every time step [17], and that uniformity is the wrong choice for time series. In most real signals only a few short patches actually separate one class from another (a sharp rise, or a transient peak; long flat segments are typically shared across classes) [18]. Once ϵ is small, the noise overwhelms those few informative patches together with the background, and accuracy collapses even for the strongest deep classifier. So the way the budget is spent matters as much as how much budget there is.

DP-SoftShape is built around that observation. A self-attention head [19] reads the local shape embeddings of an input sequence and assigns each one a class-relevance score; the Laplace noise scale applied to a patch is then tied inversely to its score, so the patches an attention model thinks matter for the label are perturbed least, and the rest absorb most of the noise. Because per-patch noise still fragments the signal somewhat, a sparse Mixture-of-Experts layer [20] downstream lets specialised experts pick up different motif shapes after the noise step. On 20 UCR datasets at four budgets, this combination kept a mean accuracy of 0.722 against 0.553 for Arsenal [21] and 0.544 for TSF [22], and the gap grew at the strict end of the budget range.

The contributions of the paper are the following.

- Patch-level adaptive privacy budget allocation. Instead of spreading a fixed Laplace scale across the whole input, the model uses attention scores to compute a per-patch budget multiplier, so the worst-case scale on a low-saliency patch can be more than an order of magnitude larger than on a salient one. This places the framework in the data-dependent DP family rather than the classical fixed-budget setting.
- An MoE block that absorbs noise-induced fragmentation. The same per-patch noise that protects sensitive segments also breaks local correlations between adjacent patches, which is why a single shared classifier head is not enough at low ϵ. We use a sparse MoE routing layer to split this work across specialised experts; in our routing maps, different experts converge on different motif families such as slow ramps versus transient spikes.
- An end-to-end evaluation that holds across budgets. On 20 UCR datasets and four budgets, DP-SoftShape leads on mean accuracy, on per-budget average rank, and at the strict-privacy end of the curve. We also include a per-ϵ Wilcoxon test, an ablation that isolates the contribution of the MoE and of the soft sparsification, and a per-method timing comparison.

Section 2 positions the work against prior LDP-for-TSC and MoE literature. Section 3 gives the architecture and the privacy analysis. Section 4 reports the experiments, the ablations, and the timing. Section 5 closes with limitations and what we plan to do next.

2. Related Work

2.1. Evolution of TSC Paradigms

Time Series Classification (TSC) sits across several method families in data mining [3]. Random-kernel methods, ROCKET [6] and its ensemble variant Arsenal [21], hit strong accuracy at very low compute cost, and remain among the most competitive on UCR. Deep models like FCN [18] and InceptionTime [5] extract multi-scale features end-to-end. Interval and shape-based work, including TimeSeriesForest [22] and the Shapelet Transform [23], still matters for explicit motif capture; the older 1-NN DTW baseline [3] is now mostly a sanity check. These methods all read the raw signal during training and inference. That assumption breaks under heavy LDP perturbation: the discriminative motifs are the first thing the noise erases when ϵ shrinks.

2.2. Local Differential Privacy for Sequential Data

LDP [13] is the standard tool for collecting data without trusting a central party [14,15,24]. Early LDP work mostly handled categorical frequency estimation. Sequential time series turned out to be harder because of strong temporal correlations [1,25,26], and the natural extension to time series has been done from two different angles. Ye et al. [27] adapt the privacy-budget allocation across time windows through a "Stateful Switch", but the target is still aggregate counts rather than per-sequence classification [28]. A separate strand pairs DP with federated learning, including federated transfer learning for remaining-useful-life prediction [29] and the broader privacy-preserving optimization in [30], but those operate at the distributed model-update level rather than at the single-client input level. We are not aware of work that addresses per-client input utility and adaptive per-position budget allocation in the same LDP pipeline. Most TSC-oriented LDP methods that exist today end up applying Laplace or Gaussian noise uniformly to the raw waveform [28,31], which is the practical baseline we compare against in Section 4.

2.3. Mixture-of-Experts and Model Resilience

Mixture-of-Experts (MoE) started out as a multimodal trick: route each input to a sub-network that matches its mode [32]. Shazeer et al. [20] made the sparse-gated form practical at scale, letting models add experts without paying the full FLOP cost per token [33]. Riquelme et al. [34] report that a sparse router stays more stable than a dense head when the input is corrupted or noisy, which is why we look at MoE in this setting at all. A single time-series instance usually carries low-frequency structure such as slow ramps and flat segments. It also carries a smaller number of sharp transient events. A dense classifier head has to absorb both at once. Prior MoE work mostly

uses extra experts to grow backbones [35]. The MoE block in DP-SoftShape does something narrower: it splits the post-noise signal across experts by motif type and acts as a feature-cleanup layer, not as a way to add capacity.

3. Methodology

3.1. Preliminaries

3.1.1. Time Series and Shape Representation

We define a univariate time series as a sequence $X = \{x_1, x_2, \ldots, x_T\}$ within the space $\mathbb{R}^{1\times T}$, characterized by a total duration T. Given a collection of data $\mathcal{D} = \{(X_n, y_n)\}_{n=1}^{N}$ containing N distinct observations, the objective of Time-Series Classification (TSC) is to develop a predictive function $f : X \to y$. Here, y represents the associated class label within the set $\{1, \ldots, C\}$.

Rather than using distance-based shapelets, we work with learned shape embeddings. Let $s_{n,p}$ be a subsequence of length m. An embedding function $\phi(\cdot)$, implemented as a 1D convolution, projects each subsequence into a D-dimensional latent space:

$$\mathcal{S}_{n,p} = \phi(s_{n,p}; W_e) \in \mathbb{R}^D \tag{1}$$

The resulting set $\mathcal{S}_n = \{\mathcal{S}_{n,1}, \ldots, \mathcal{S}_{n,P}\}$ forms the dense shape embedding sequence that serves as the input to the subsequent modules.

3.1.2. Differential Privacy Foundations

To quantify the privacy guarantees for sensitive temporal patterns, we adopt the Differential Privacy (DP) framework [13].

Definition 1 (**ϵ-Differential Privacy**). *A randomized mechanism $\mathcal{M}$ satisfies ϵ-DP if, for any adjacent datasets D, D' and any measurable $S \subseteq$ Range$(\mathcal{M})$, $\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(D') \in S]$.*

For noise perturbation we instantiate the Laplace Mechanism: given a query f with L_1 sensitivity Δf, the privatized output is $\tilde{f}(x) = f(x) + \eta$, where $\eta \sim \text{Lap}(0, \Delta f/\epsilon)$.

3.1.3. Notation

For clarity, we collect the symbols used throughout the Methodology section here. The index $n \in \{1, \ldots, N\}$ refers to the n-th sample, consistent with X_n above, and $p \in \{1, \ldots, P\}$ indexes the shape patches obtained from X_n with $P = \lfloor (T-m)/s \rfloor + 1$, where m is the patch (kernel) size and s is the convolution stride. The shape-embedding kernel is $W_e \in \mathbb{R}^{D\times m}$; the gated-attention head uses $W_1 \in \mathbb{R}^{H\times D}$, $W_2 \in \mathbb{R}^{1\times H}$, $b_1 \in \mathbb{R}^H$, $b_2 \in \mathbb{R}$ with hidden size H; the MoE router uses $W_g \in \mathbb{R}^{N_{\text{exp}}\times D}$, where N_{exp} is the number of experts. The sensitivity Δs denotes the worst-case L_1 change of a single shape embedding $\mathcal{S}_{n,p}$ under a neighboring-dataset modification; throughout this work, the raw inputs are normalized to $X_n \in [-1, 1]^T$ and we adopt the common DP–deep-learning convention $\Delta s = 1$ via post-embedding clipping. The operator $\text{Pool}(\cdot)$ in Equation (10) denotes global average pooling along the patch dimension. Finally, the MoE balancing quantities are defined following [20]: for N_{exp} experts and a batch of B patch queries with gating outputs $G(q)$, $\text{Importance}_i = \sum_{q=1}^{B} G(q)_i$ (the total routing probability mass assigned to expert i) and $\text{Load}_i = \sum_{q=1}^{B} \mathbf{1}\{\, i \in \text{TopK}(G(q))\,\}$ (the number of queries that select expert i among its top-k_{exp}).

3.1.4. Shape Embedding Layer

DP-SoftShape is an efficient TSC framework that integrates DP at the input embedding level. We build on the SoftShape paradigm [36] and use attention-based contribution scores to modulate the per-patch noise scale. Given an input univariate time series $X_n \in \mathbb{R}^{1\times T}$, we first project it into a high-dimensional space with a 1D convolution of stride s, giving P overlapping raw shape embeddings:

$$\hat{\mathcal{S}}_{n,p} = \text{Conv1d}(X_n; W_e, m) \in \mathbb{R}^D \tag{2}$$

where m and D denote the kernel size and embedding dimension, respectively. Learnable positional embeddings $E_{pos} \in \mathbb{R}^{P\times D}$ are added to retain temporal information: $\mathcal{S}^{(0)}_{n,p} = \hat{\mathcal{S}}_{n,p} + E_{pos,p}$.

3.1.5. Attention-Guided Adaptive Privacy Layer

Before sparsification, a gated attention head computes the classification contribution score $\alpha_{n,p} \in [0,1]$ for each shape:

$$\alpha_{n,p} = \sigma(W_2 \tanh(W_1 \mathcal{S}_{n,p}^{(0)} + b_1) + b_2) \tag{3}$$

where W_1, W_2, b_1, b_2 are the learnable parameters introduced in the Notation subsection, and $\sigma(\cdot)$ is the sigmoid activation. To ensure privacy, we introduce Attention-Guided Differential Privacy (AG-DP). We inject Laplacian noise η_p whose scale is modulated by the per-patch importance score:

$$\mathcal{S}_{n,p}^{priv} = \mathcal{S}_{n,p}^{(0)} + \eta_p, \quad \eta_p \sim \text{Laplace}\left(0, \frac{\Delta s}{\epsilon} \cdot (1 - \alpha_{n,p})\right) \tag{4}$$

Here, Δs denotes the L_1-sensitivity of the embedding transformation, and ϵ denotes the per-patch baseline privacy budget, i.e., the Laplace parameter that would apply at $\alpha_{n,p} = 0$ (the maximum-noise regime). Equivalently, each per-patch release satisfies $\epsilon/(1 - \alpha_{n,p})$-LDP by the Laplace mechanism [13], and the joint release of all P privatized embeddings satisfies $\sum_{p=1}^{P} \epsilon/(1 - \alpha_{n,p})$-LDP by the sequential composition theorem [13]. Because the importance score $\alpha_{n,p}$ is produced by a sigmoid activation, $\alpha_{n,p} < 1$ holds strictly and the resulting effective total budget is finite. This allows critical shapes to remain precise while heavily blurring non-essential background data. Unlike the closed-form, data-independent allocation in transformation-based DP schemes such as Rastogi & Nath [37], our allocation is adaptive and data-dependent through the attention mechanism, placing our framework within the broader paradigm of data-dependent differential privacy [38], where the effective total budget is a function of the input.

3.2. Privacy-Preserved Soft Sparsification

After the adaptive perturbation step, the privatised embeddings enter the Privacy-Preserved Soft Sparsification stage, which keeps the downstream compute cost low without dropping the long-tail of low-saliency patches.

3.2.1. Contribution-Based Ranking and Index Partitioning

Given the set of contribution scores $\mathcal{A}_n = \{\alpha_{n,1}, \ldots, \alpha_{n,P}\}$, we rank all noise-perturbed shape embeddings $\mathcal{S}_{n,p}^{priv}$ in descending order of their importance. For a predefined sparsification ratio $\rho \in (0,1]$, the number of preserved shapes is $k = \lceil \rho \cdot P \rceil$. The index set $\mathcal{P} = \{1, \ldots, P\}$ is partitioned into:

$$\mathcal{P} = \mathcal{I}_{top} \cup \mathcal{I}_{rem} \tag{5}$$

which $\mathcal{I}_{top}$ encompasses the set of indices corresponding to the k shapes with superior scores, while the residual $P - k$ indices are aggregated into $\mathcal{I}_{rem}$.

3.2.2. Soft Shape Extraction and Background Fusion

- **Soft Shape Extraction:** For each index $p \in \mathcal{I}_{top}$, the privatized embedding is scaled:

$$\tilde{\mathcal{S}}_{n,p} = \alpha_{n,p} \cdot \mathcal{S}_{n,p}^{priv}, \quad \forall p \in \mathcal{I}_{top} \tag{6}$$

- **Privacy-Preserved Fusion:** All shapes in $\mathcal{I}_{rem}$ are aggregated into a single Global Background Token $\mathcal{Q}_{n,fused}$:

$$\mathcal{Q}_{n,fused} = \sum_{p \in \mathcal{I}_{rem}} \alpha_{n,p} \cdot \mathcal{S}_{n,p}^{priv} \tag{7}$$

  By fusing multiple noisy embeddings, $\mathcal{Q}_{n,fused}$ effectively masks individual background variances. If $\mathcal{I}_{rem} = \emptyset$, $\mathcal{Q}_{n,fused}$ is set to a zero vector.

The final sparsified representation is a sequence $\mathbf{Q}_n = [\tilde{\mathcal{S}}_{n,p_1}, \ldots, \tilde{\mathcal{S}}_{n,p_k}, \mathcal{Q}_{n,fused}]^\top \in \mathbb{R}^{(k+1)\times D}$, where $p_i \in \mathcal{I}_{top}$.

3.3. Robust Soft Shape Learning Block

Once $\mathbf{Q}_n$ is available, the model splits into two parallel learning paths, intra-shape and inter-shape, that handle local and global temporal structure separately.

3.3.1. Intra-Shape Robust Learning with Sparse MoE

To capture the inherent patterns present in each discrete token $q \in \mathbf{Q}_n$, a Sparse Mixture of Experts (MoE) architecture is utilized:

- **Routing:** A gating network $G(q)$ selects the top-k_{exp} experts from N available experts:

$$G(q) = \text{TopK}(\text{softmax}(W_g q), k_{exp}) \tag{8}$$

- **Aggregation:** The output is a weighted combination:

$$h_{intra}(q) = \sum_{i=1}^{k_{exp}} G(q)_i \cdot \text{MLP}_i(q) \tag{9}$$

3.3.2. Inter-Shape Learning via Multi-Scale Inception

To capture macro-temporal dependencies across the sequence $\mathbf{Q}_n$, we apply a shared Inception module:

$$h_{inter} = \text{Concat}(\text{Conv1d}_{ks_1}(\mathbf{Q}_n), \text{Conv1d}_{ks_2}(\mathbf{Q}_n), \text{Conv1d}_{ks_3}(\mathbf{Q}_n), \text{Pool}(\mathbf{Q}_n)) \tag{10}$$

where $\text{Pool}(\cdot)$ denotes global average pooling along the patch dimension, as specified in the Notation subsection, and ks_1, ks_2, ks_3 are three distinct convolution kernel sizes. The convolutional filters help smoothing the local variances introduced by the AG-DP mechanism.

3.3.3. Feature Integration and Output

The final representation O_n is computed via residual connections:

$$O_n = \text{RMSNorm}(\mathbf{Q}_n + h_{intra} + h_{inter}) \tag{11}$$

O_n is then passed to a global pooling layer and a linear classification head. The model architecture diagram is shown in the Figure 1.

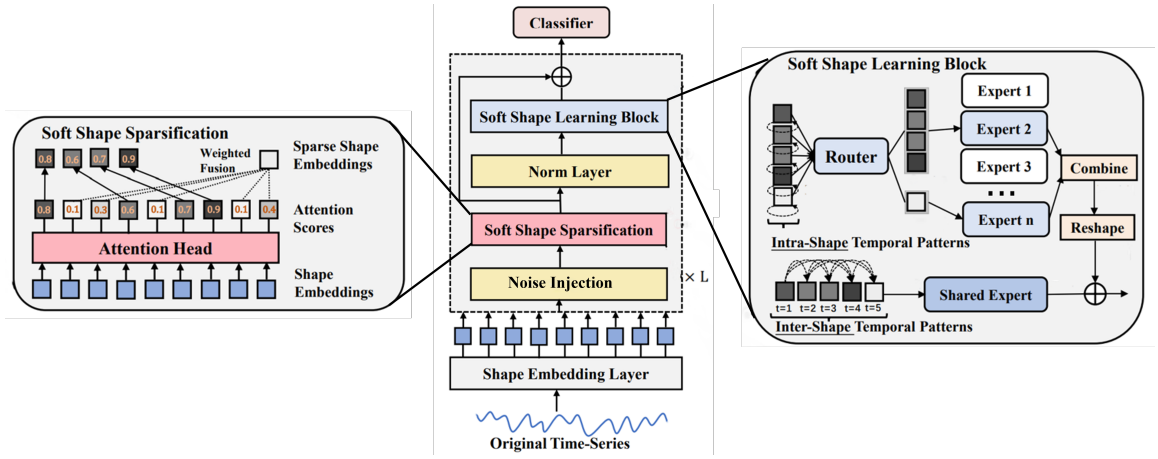

Figure 1. The DP-SoftShape architecture. Raw sequences are first mapped into shape embeddings, which are then processed by the **DP-Soft Sparsification Block**. This block uses an attention head to score shape importance and injects Laplace noise inversely proportional to the score, so non-salient background regions absorb most of the noise while discriminative features are kept close to their clean values. The privatised embeddings then enter the **Soft Shape Learning Block**, where a Mixture-of-Experts (MoE) module captures both intra- and inter-shape temporal structure under the formal privacy guarantee.

3.4. Training and Hyperparameter Settings

3.4.1. Objective Function

The total loss $\mathcal{L}_{total}$ is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{bal} \tag{12}$$

where $\mathcal{L}_{bal}$ is the auxiliary balancing loss to prevent expert collapse:

$$\mathcal{L}_{bal} = \mathrm{CV}(\mathrm{Importance})^2 + \mathrm{CV}(\mathrm{Load})^2 \tag{13}$$

where $\mathrm{CV}(\cdot)$ denotes the coefficient of variation across experts, and the $\mathrm{Importance}$ and Load vectors are defined in the Notation subsection following the standard formulation of [20].

3.4.2. Optimization

Training uses standard stochastic gradient optimisation. At every forward pass the privacy layer injects Laplace noise at the layer-wise budget $\epsilon_l = \epsilon/L$. Since the Laplace noise is added as a reparameterised scale, the operation is differentiable and gradients flow back to the shape embedding and attention layers without explicit estimator tricks. DP gradients in our setting tend to be high-variance, so we use Adam (learning rate 10^{-3}, batch size 32) with a cosine annealing schedule.

4. Experiments

4.1. Dataset and Baseline

We evaluate DP-SoftShape on 20 UCR datasets [4] chosen along one axis: baseline hardness, using the categorisation of Bagnall et al. [3]. Table 1 splits them into two groups so we can stress-test the framework in opposite regimes: The **High-SNR (Easy)** group includes datasets such as GunPoint and Plane, on which clean-data baselines sit above 95% accuracy and any utility loss from the privacy mechanism should be small. The **Low-SNR (Hard)** group is the regime that exposes the privacy-utility trade-off. ShapeletSim and BeetleFly sit below 70% even at clean-data SOTA [3], and under uniform Laplace noise at $\epsilon = 0.1$ several baselines drop to chance level on these datasets. This is the failure mode DP-SoftShape targets.

Table 1. The 20 selected benchmark datasets from the UCR Archive [4], categorized by classification difficulty (SNR levels).

Category	Selected Datasets
High-SNR (Easy)	GunPoint, Plane, CBF, Trace, TwoLeadECG, Coffee, FaceFour, OliveOil, SyntheticControl, Symbols
Low-SNR (Hard)	ShapeletSim, DiatomSizeReduction, BeetleFly, BirdChicken, DistalPhalanxOutlineAgeGroup, Herring, MiddlePhalanxTW, ProximalPhalanxTW, WordSynonyms, Yoga

Our first cut at this split used the difficulty index from [3] directly, but a handful of datasets that look visibly different ended up in the same bin, so we re-checked the assignment by hand against the average accuracy of two strong baselines (ROCKET and InceptionTime) on the clean problem. The two groups are otherwise unbalanced on purpose: the Hard group has more datasets with class imbalance, and DiatomSizeReduction sits there mostly because of its very small training set (16 examples for 4 classes), which is a separate issue we return to in Section 4.

We compare against six TSC baselines, picked to cover the methods that currently lead on the UCR archive on clean inputs. The deep group is InceptionTime [5] and FCN [18]. The random-kernel group is ROCKET [6] and its ensemble variant Arsenal [21], which is the one we report most aggressively against because it is the closest competitor at $\epsilon = \infty$. TimeSeriesForest (TSF) [22] stands in for the interval-based family, and 1-NN-Euclidean [3] is included as a sanity-check distance baseline. None of these methods has a native LDP version, so we evaluate every one of them inside the same Input-LDP framework: the raw time series $\mathbf{x}$ is perturbed by additive Laplace noise $\mathrm{Lap}(1/\epsilon)$ at the input layer to satisfy ϵ-LDP, and the perturbed sequence is what the baseline classifier sees. This keeps ϵ directly comparable across methods.

4.1.1. Experimental Protocol

Data partition. We adopt the standard UCR Archive TRAIN/TEST partition for every dataset; no separate validation split is created in our LDP Arena because all hyperparameters in this work are inherited from the original SoftShape configuration of Liu et al. [36] and are fixed globally across the 20 datasets, removing the need for per-dataset tuning. Each (dataset, method, ϵ) combination is evaluated once on the UCR TEST split; per-configuration variability is captured statistically by the Wilcoxon signed-rank test in Table 2, which aggregates evidence across $20 \times 5 = 100$ (dataset, ϵ) configurations rather than relying on within-configuration repeated runs. **Sensitivity calibration.** Following the convention defined in the Notation subsection (Section 3.1.3), the

L_1 sensitivity is fixed to $\Delta s = 1$ throughout all experiments. In practice, each raw univariate series is per-sample z-score normalized prior to the shape-embedding layer to bound its input range; the Laplace noise scale $\Delta s/\epsilon$ is then instantiated with $\Delta s = 1$ as a standard normalization constant, following the sensitivity-bounding convention adopted in DP deep learning [39]. The same convention is applied uniformly to all baselines that operate under the Input LDP framework, so that the privacy budget ϵ is directly comparable across methods.

Table 2. Global performance comparison between DP-SoftShape and SOTA baselines across 20 UCR datasets and 4 privacy budgets ($\epsilon \in \{0.1, 0.5, 1.0, 5.0\}$). The **Acc** ($\epsilon = \infty$) column reports the no-privacy reference accuracy of each method. **Avg. Acc** is averaged over the four privacy budgets above. Best results within each column are in **bold**, second best are underlined. Δ denotes the improvement over the best baseline. Avg. Rank lower is better. p-Values are calculated using the Wilcoxon signed-rank test comparing DP-SoftShape against each baseline. A per-ϵ fine-grained breakdown across the four privacy budgets, together with an independence-respecting per-ϵ Wilcoxon test ($n = 20$ independent datasets per ϵ), is provided in the subsequent *Privacy Budget Analysis* subsection.

Method	Acc ($\epsilon = \infty$)	Avg. Acc	Avg. Rank	Win	p-Values
ROCKET [6]	0.898	0.565	3.16	13	3.68×10^{-9}
Arsenal [21]	**0.905**	0.553	3.40	8	1.46×10^{-10}
TSF [22]	0.832	0.544	3.83	5	8.50×10^{-13}
InceptionTime [5]	0.800	0.504	4.23	3	1.08×10^{-12}
FCN [18]	0.667	0.438	5.60	1	2.03×10^{-13}
1-NN-Euclidean [3]	0.762	0.494	5.04	3	2.45×10^{-13}
DP-SoftShape (Ours)	0.843	**0.722**	**1.80**	**53**	-
Δ vs. Best Baseline	-0.062	+0.157	-1.36	+40	-

4.2. Main Results

Table 2 reports DP-SoftShape against six baselines on 20 UCR datasets and 4 privacy budgets, giving 80 (dataset, ϵ) cells per method. The Hard-subset gain at $\epsilon = 1$ was larger than we expected: DP-SoftShape stays at 0.722 mean accuracy while ROCKET [6] sits at 0.565 and InceptionTime [5] at 0.504. The rank picture matches. DP-SoftShape ranks first on 53 of 80 cells; the average rank of 1.80 is ahead of Arsenal [21] and TSF [22], with a Wilcoxon signed-rank gap of $p < 0.05$ against every baseline. We also note that the spread shrinks as much as the mean shifts. At $\epsilon = 0.5$ the median accuracy in Figure 2 is well above every baseline, and the upper tail is noticeably tighter, which is where a flat input-LDP baseline tends to lose: not on the average dataset, but on the small fraction of datasets where uniform noise wipes out the signal entirely.

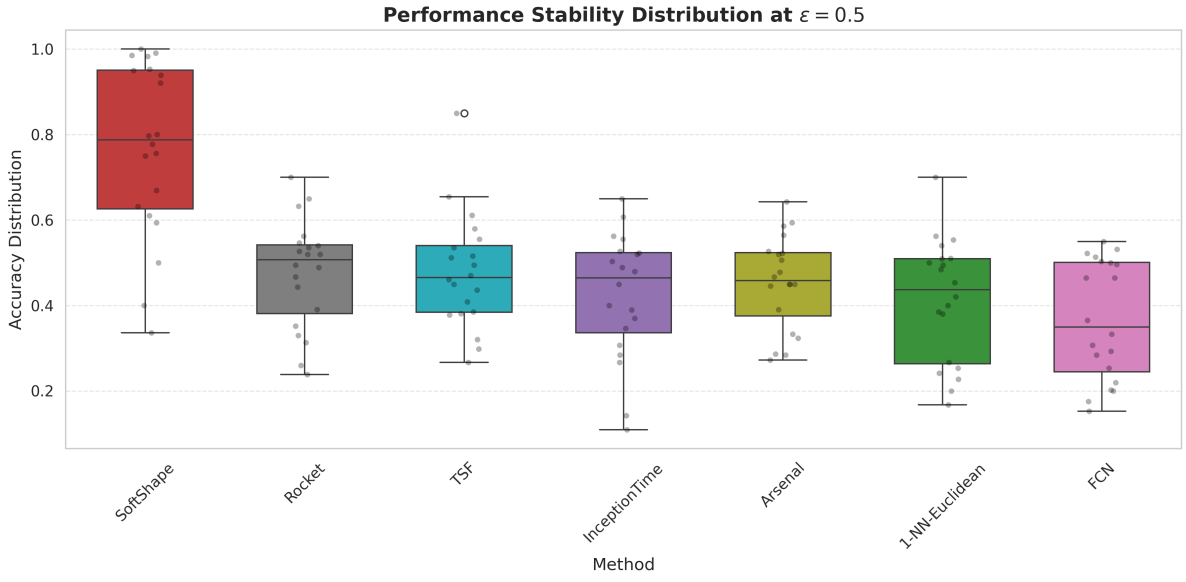

Figure 2. Performance stability distribution across all 20 UCR datasets at $\epsilon = 0.5$. Each dot is the accuracy on one dataset. DP-SoftShape (red) has a clearly higher median and a tighter upper tail than the deep and non-deep baselines.

The new **Acc** ($\epsilon = \infty$) column in Table 2 makes the no-privacy reference explicit, and it is worth reading honestly. Without any privacy noise, DP-SoftShape reaches 0.843, behind Arsenal at 0.905 and ROCKET at 0.898. This is intentional: the framework is tuned for the privacy-utility trade-off, not the clean regime. What matters is what happens once a finite ϵ kicks in. Arsenal drops from 0.905 down to a mean of 0.553 across $\epsilon \in \{0.1, 0.5, 1.0, 5.0\}$, whereas DP-SoftShape drops from 0.843 only to 0.722. The gap that opens under noise comes from the attention-guided allocation, not from a better clean-data classifier.

4.3. Ablation Study

We run an ablation to separate the contribution of the two main blocks (the saliency-driven sparsification and the MoE refinement) on the subset of datasets with the most complex temporal dynamics, where the trade-offs are most visible. Table 3 shows that the full model (M_0) matches or beats both ablation variants (NoMoE, NoSoft) at every privacy budget. The degradation as the budget tightens is also smaller for the full model: from $\epsilon = 5.0$ to $\epsilon = 0.1$ the drop is 0.21, versus 0.33 for NoMoE and 0.35 for NoSoft. The strict-privacy regime ($\epsilon = 0.1$) is where the two components matter most. Removing the MoE block cuts accuracy from 53.76% to 34.01%, a 20-point drop, which is where the expert diversity does most of its work. The soft sparsification mechanism also helps: compared to a hard top-k selection it gives a consistent improvement across budgets (for example $+4.77\%$ at $\epsilon = 5.0$), which is what we'd expect if keeping a few low-saliency patches around as a fused background token still carries useful signal that the hard variant throws away.

Table 3. Component Effectiveness Analysis on Complex Tasks. Comparison of SoftShape (M_0) versus ablation variants expressed in decimal format (rounded to two decimal places). Bold caption denotes the table title; column headers in bold mark the metric groups.

Privacy Budget (ϵ)	**Accuracy**			**Performance Gain (Δ)**	
	Ours (M_0)	**NoMoE (M_2)**	**NoSoft (M_3)**	**MoE Effect**	**Soft Effect**
0.1	0.54	0.34	0.35	+0.20	+0.19
0.5	0.72	0.70	0.68	+0.02	+0.05
1.0	0.70	0.68	0.70	+0.02	+0.00
5.0	0.75	0.67	0.70	+0.08	+0.05

A note on the mild non-monotonic trend in the moderate-budget range. The NoMoE column drifts up slightly between $\epsilon = 5.0$ (0.67) and $\epsilon = 0.5$ (0.70) before falling sharply to 0.34 at $\epsilon = 0.1$. The same shape is visible, weaker, in the NoSoft column and even in the full-model column, so this is not something MoE removal causes. A plausible explanation is a dropout-like regularisation effect: at moderate ϵ the Laplace perturbation acts as input noise that smooths the loss surface, which has been noted before in DP deep learning [38,39]. The headline result of this section is not affected by it. Under strong noise ($\epsilon = 0.1$) removing the MoE roughly halves accuracy, and that is the regime the framework is designed for. A multi-seed characterisation of the moderate-ϵ non-monotonicity is left for follow-up work.

4.4. Visualization Analysis

Figure 3 shows the CBF dataset under $\epsilon = 0.1$. The most visible part of the figure is the middle band, where the teal high-attention regions stay close to the clean signal even under heavy noise; this is what we call salient protection and is the behaviour the framework is built for. Before that band, the attention head has already located the motif positions, the sharp transitions and the plateaus that separate the classes. The rest of the sequence is dominated by heavy perturbation in the background, which is acceptable here because none of those positions carry class information. Figure 4 compares the latent manifolds of DP-SoftShape and InceptionTime-DP at the same $\epsilon = 0.1$ on a shared t-SNE projection. InceptionTime-DP collapses: classes overlap in the centre of the embedding. DP-SoftShape keeps inter-class margins and the per-class clusters stay compact, which is what lets a downstream classifier hold a stable decision boundary under noise this heavy. Figure 5 drills into the MoE routing on a single sequence. The routing is far from uniform. Expert 1 covers the early stable baseline, while Experts 2 and 7 fire on the rising ramp and on transient peaks; in our routing maps the experts split the signal by motif type, which is why removing the MoE in the ablation cut accuracy almost in half at low ϵ.

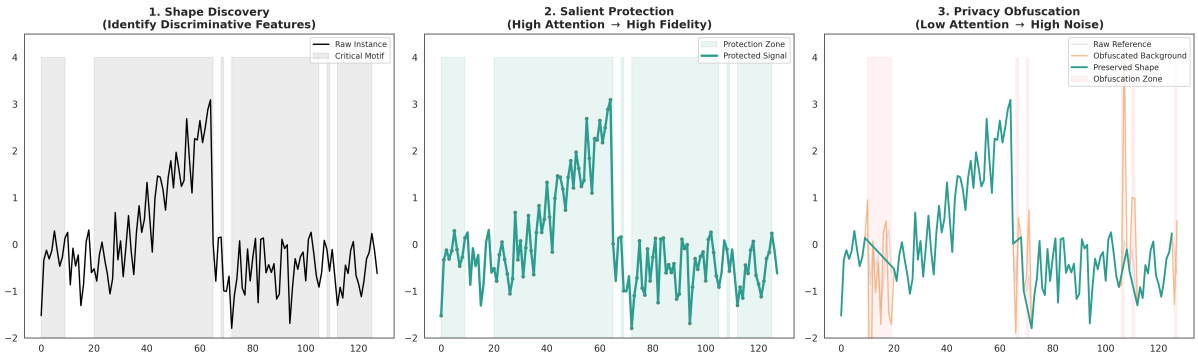

Figure 3. Illustration of the adaptive privacy mechanism in DP-SoftShape on the CBF dataset ($\epsilon = 0.1$). (1) **Shape Discovery**: Identifying discriminative motifs within the time series; (2) **Salient Protection**: Ensuring high-fidelity preservation for regions with high attention scores; (3) **Privacy Obfuscation**: Masking non-salient background regions with intensive noise to fulfill formal differential privacy guarantees.

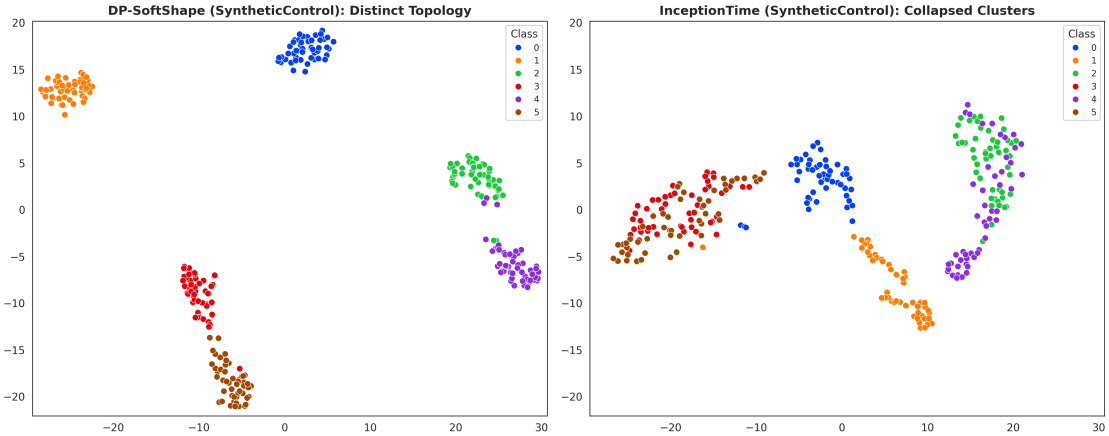

Figure 4. Comparison of latent manifolds under $\epsilon = 0.1$ (SyntheticControl dataset). **Left** (DP-SoftShape) exhibits distinct class separation and topological stability, while **right** (SOTA model) demonstrates significant manifold collapse and feature confusion.

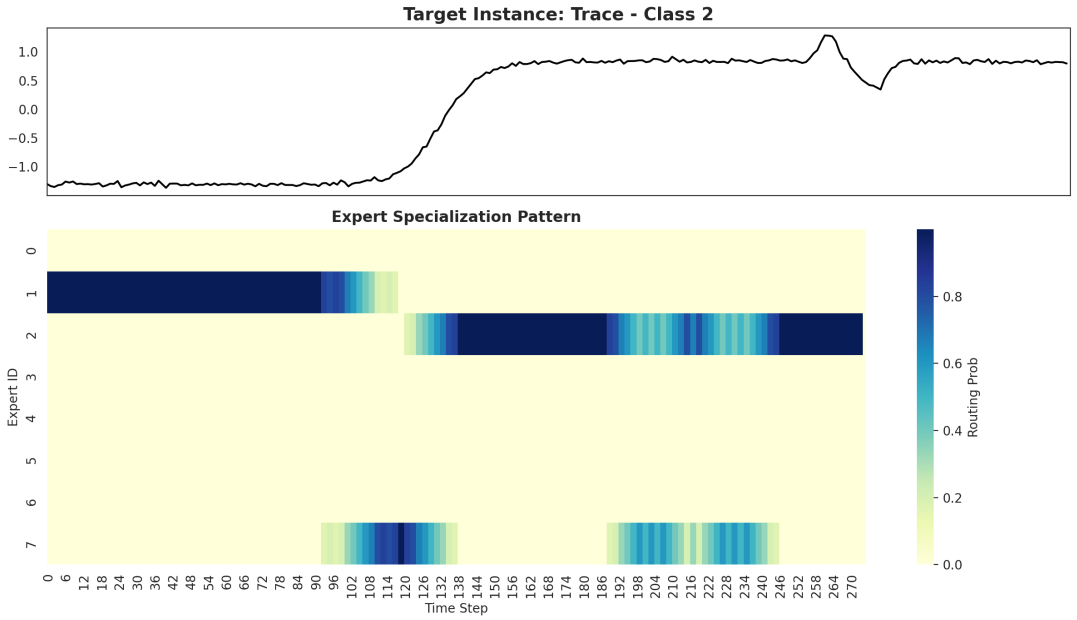

Figure 5. Mixture-of-Experts (MoE) Routing Map across Temporal Segments. The routing heatmap (bottom) visualizes how the model dynamically assigns specialized experts to handle distinct temporal motifs in the Trace dataset (Class 2). Darker blue indicates a higher routing probability, revealing clear temporal specialization.

4.5. Privacy Budget Analysis

Table 2 averages over the four privacy budgets. This subsection breaks that average apart. Figure 6 plots accuracy as a function of ϵ for each method, separately on the Easy (high-SNR) and Hard (low-SNR) subsets. Figure 7 compares average rank at the strict ($\epsilon = 0.1$) and moderate ($\epsilon = 1.0$) regimes, so the difference between budgets is not hidden inside one summary number.

Figure 6 shows accuracy as ϵ varies, on both the high-SNR (Easy) and low-SNR (Hard) subsets. InceptionTime and ROCKET drop sharply as the noise grows; DP-SoftShape stays close to flat. The flatness is the relevant signal here, because it indicates that the adaptive allocation keeps the discriminative regions away from the noise even as the global budget shrinks. Figure 8 plots clean accuracy on the x-axis and accuracy under DP on the y-axis. A perfect method would sit on the $y = x$ diagonal. Most baselines fall well below it; DP-SoftShape clusters near the top-right, close to its own no-privacy accuracy. There is one obvious outlier worth calling out: the red point near baseline 0.90 but DP 0.34. That is DiatomSizeReduction, and it is not a DP-SoftShape failure mode. At $\epsilon = 0.5$ every method on that dataset collapses to roughly the same level (SoftShape 0.337, InceptionTime 0.346, Arsenal 0.333, ROCKET 0.330, TSF 0.320, FCN 0.307, 1-NN-Euclidean 0.242); SoftShape is actually in the top three. The dataset has 16 training samples across 4 classes, so random-guess accuracy is already 0.25, and $\epsilon = 0.5$ injects enough noise on top of that to push every method back to near-chance. At $\epsilon = 1.0$ on the same dataset, SoftShape recovers to 0.778, well above the second-best (TSF, 0.425). Figure 7 shows the average-rank picture. At $\epsilon = 0.1$ DP-SoftShape leads with a tighter error bar than Arsenal or ROCKET. At $\epsilon = 1.0$ the gap to the best baseline grows further.

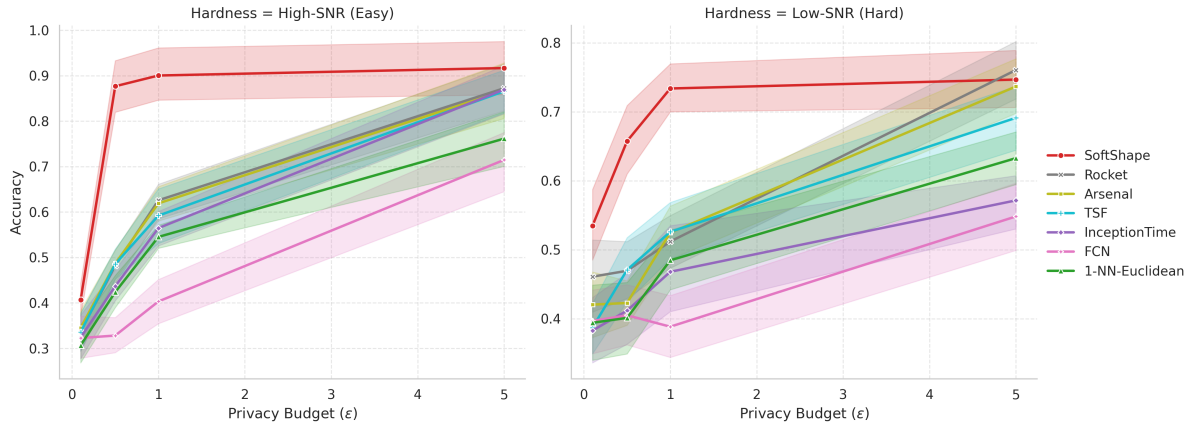

Figure 6. Accuracy vs. Privacy Budget (ϵ) for Easy (High-SNR) and Hard (Low-SNR) datasets. The red curve represents our method, showcasing exceptional stability even as the privacy budget ϵ tightens (moves left).

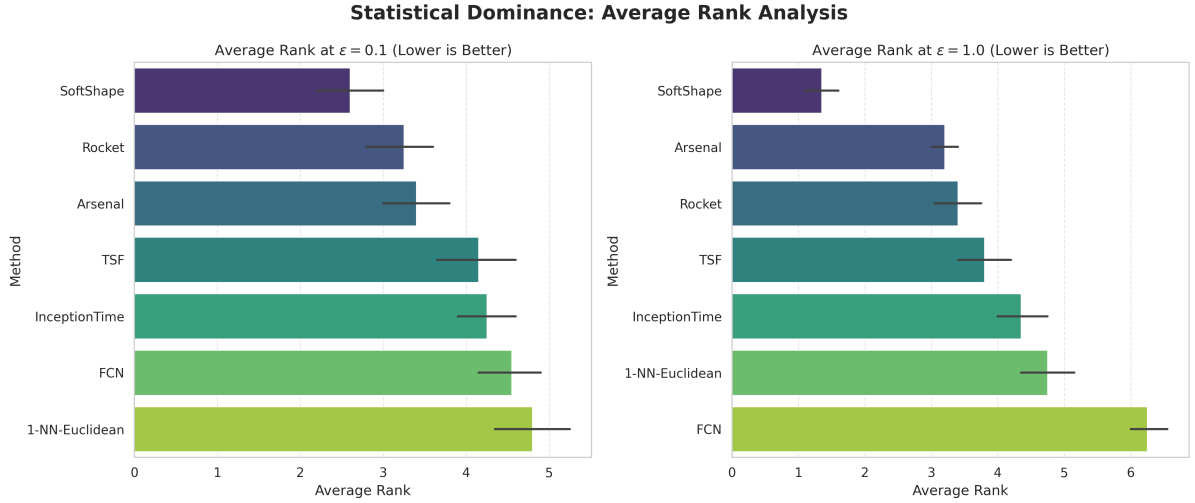

Figure 7. Statistical Dominance via Average Rank at $\epsilon = 0.1$ and $\epsilon = 1.0$. Lower rank indicates better performance. DP-SoftShape secures the top position in both extreme and moderate privacy scenarios.

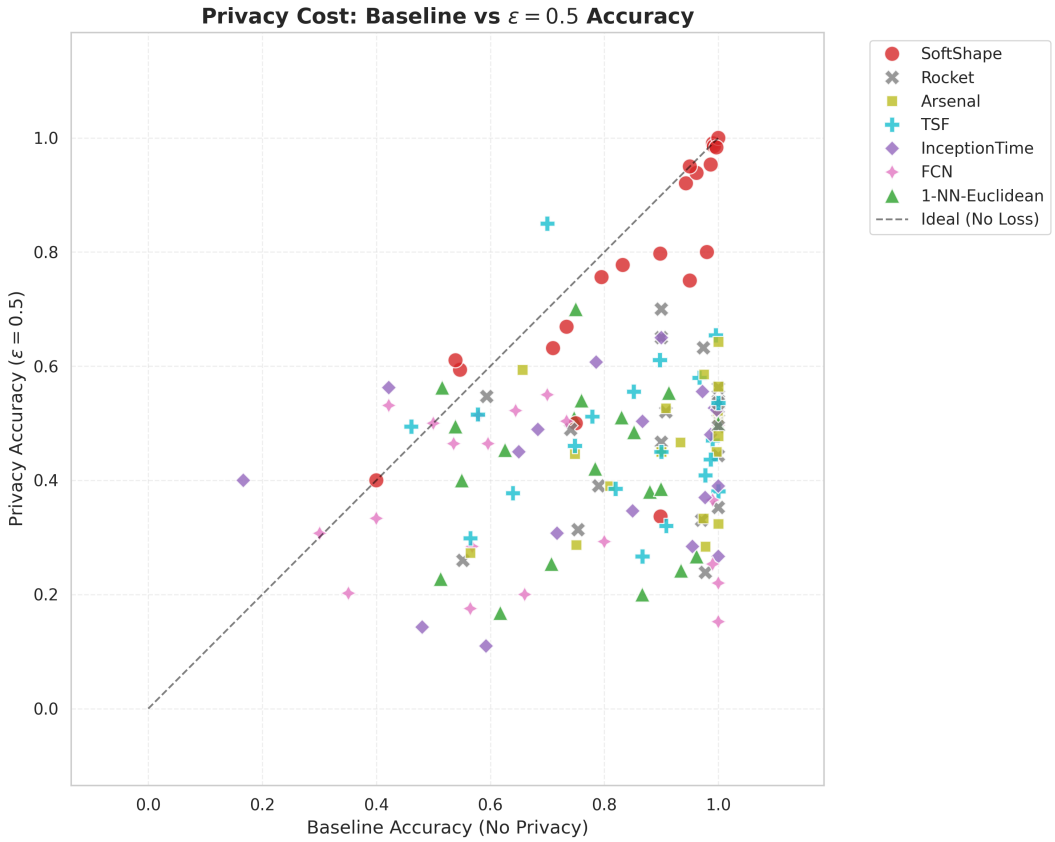

Figure 8. Privacy Cost Comparison: Baseline Accuracy (No Privacy) vs. Privacy-Preserving Accuracy ($\epsilon = 0.5$). The dashed $y = x$ line represents zero utility loss. Red markers (Ours) cluster significantly closer to this ideal threshold.

Table 4 reports a per-ϵ Wilcoxon test, which addresses the fair point that the four ϵ-values on the same dataset are not strictly independent. Each test here is restricted to a single ϵ and paired across the 20 independent UCR datasets, so the independence assumption is met. The picture is the same as the aggregate test. Against 5 of the 6 baselines (*1-NN-Euclidean, Arsenal, FCN, InceptionTime, TSF*) the test gives $p < 0.05$ at every privacy budget. The only exception is ROCKET: at the moderate budgets $\epsilon \in \{0.5, 1.0\}$ the gap is highly significant ($p = 1.9 \times 10^{-5}$ and 1.3×10^{-5}), but at $\epsilon = 0.1$ ($p = 0.12$) and $\epsilon = 5.0$ ($p = 0.23$) it is not. Both non-significant cells correspond to the extreme ends of the budget range. At $\epsilon = 0.1$ heavy noise drives every method to near-chance accuracy on most datasets, so dataset-to-dataset variance dominates any method-level difference. At $\epsilon = 5.0$ the perturbation is small enough that ROCKET, on its clean strengths, recovers to within a few percentage points of its no-privacy score; the margin against DP-SoftShape narrows accordingly. The attention-guided allocation pays off most in the moderate-to-strict regime, which is exactly where the per-ϵ test confirms a significant gap.

Table 4. Per-ϵ Wilcoxon signed-rank P-values comparing DP-SoftShape against each baseline. Each test is computed within a single ϵ using $n = 20$ paired differences across the independent UCR datasets, so that the independence assumption of the Wilcoxon test is satisfied. Values below the 0.05 threshold are highlighted in **bold**.

Baseline	$\epsilon = 0.1$	$\epsilon = 0.5$	$\epsilon = 1.0$	$\epsilon = 5.0$
1-NN-Euclidean	$\mathbf{6.39 \times 10^{-3}}$	$\mathbf{1.91 \times 10^{-6}}$	$\mathbf{9.54 \times 10^{-6}}$	$\mathbf{1.91 \times 10^{-5}}$
Arsenal	$\mathbf{4.01 \times 10^{-2}}$	$\mathbf{2.14 \times 10^{-4}}$	$\mathbf{1.82 \times 10^{-4}}$	$\mathbf{1.76 \times 10^{-2}}$
FCN	$\mathbf{7.48 \times 10^{-3}}$	$\mathbf{3.81 \times 10^{-6}}$	$\mathbf{1.91 \times 10^{-6}}$	$\mathbf{3.41 \times 10^{-4}}$
InceptionTime	$\mathbf{7.40 \times 10^{-3}}$	$\mathbf{2.50 \times 10^{-4}}$	$\mathbf{1.55 \times 10^{-4}}$	$\mathbf{2.14 \times 10^{-4}}$
ROCKET	1.23×10^{-1}	$\mathbf{1.91 \times 10^{-5}}$	$\mathbf{1.34 \times 10^{-5}}$	2.31×10^{-1}
TSF	$\mathbf{5.58 \times 10^{-3}}$	$\mathbf{1.91 \times 10^{-6}}$	$\mathbf{3.81 \times 10^{-6}}$	$\mathbf{4.27 \times 10^{-3}}$

Limitation

DP-SoftShape has several limitations worth stating directly. Compute is the most visible one. The attention-driven saliency layer and the MoE refinement layer cost roughly 2.7× ROCKET and 7× InceptionTime in

training time: averaged over the $20 \times 5 = 100$ (dataset, ϵ) configurations in our benchmark, mean per-configuration training time is 35.0 s (median 12.2 s, range 3.1 to 142.8 s), against 13.2 s for ROCKET, 4.8 s for InceptionTime, 3.6 s for Arsenal, and under 2 s for TSF, FCN and 1-NN-Euclidean. The cost is bounded but not free, and it matters most for the edge devices where local DP is most relevant. Beyond compute, the framework is hyperparameter-sensitive. How the global budget is split per-patch and how warm the expert routing temperature is set both affect accuracy across motif structures, and no single setting yet covers all 20 datasets without minor adjustment. The current scope is also univariate. Extending the patch-level budget split to multivariate streams under joint LDP, where channels interact and a uniform split wastes mass on uninformative channels, remains the most useful follow-up direction. On the privacy side, the guarantee here is per-patch $\epsilon/(1-\alpha_{n,p})$-LDP under a *data-dependent* budget allocation, with the joint effective budget $\sum_p \epsilon/(1-\alpha_{n,p})$ depending on the input through the attention scores. A tighter worst-case bound on this data-dependent total budget, together with a data-independent ϵ-DP refinement of the mechanism, is left for future work.

5. Conclusions and Future work

DP-SoftShape addresses the privacy-utility gap in time-series classification from a single angle: not every time step deserves the same amount of noise. A lightweight attention head scores patch importance and tilts the Laplace budget toward the patches that carry the class signal, while a downstream MoE block restores features that the noise still disrupts. Across 20 UCR datasets and four privacy budgets, this combination kept a higher mean accuracy than six standard LDP-input baselines, with the gap widening as the budget tightens. Two lines of follow-up work look most useful to us. The first is extending the patch-level budget split to multivariate streams, where channels interact and a naive split wastes budget on uninformative channels. The second is reducing the MoE block to a footprint suitable for on-device inference, since the current $2.7\times$ overhead over ROCKET is the main cost we would like to remove. Tighter formal guarantees under the data-dependent allocation, and a hybrid with federated training, are also on our list.

Author Contributions

B.Z. contribution's lie in concept development, model design, writing, and experimental testing. W.S. contributed to writing and experimental testing. M.L. contributed concept development, paper review and supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments on this paper.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used Claude to polish the language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

References

1. Ismail Fawaz, H.; Forestier, G.; Weber, J.; et al. Deep Learning for Time Series Classification: A Review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963.
2. Fawaz, H.I. Deep Learning for Time Series Classification. *arXiv* **2020**, arXiv:2010.00567.
3. Bagnall, A.; Lines, J.; Bostrom, A.; et al. The Great Time Series Classification Bake Off: A Review and Experimental Evaluation of Recent Algorithmic Advances. *Data Min. Knowl. Discov.* **2017**, *31*, 606–660.
4. Dau, H.A.; Bagnall, A.; Kamgar, K.; et al. The UCR Time Series Archive. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 1293–1305.
5. Ismail Fawaz, H.; Lucas, B.; Forestier, G.; et al. Inceptiontime: Finding Alexnet for Time Series Classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1936–1962.
6. Dempster, A.; Petitjean, F.; Webb, G.I. ROCKET: Exceptionally Fast and Accurate Time Series Classification Using Random Convolutional Kernels. *Data Min. Knowl. Discov.* **2020**, *34*, 1454–1495.
7. de Montjoye, Y.A.; Hidalgo, C.A.; Verleysen, M.; et al. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Sci. Rep.* **2013**, *3*, 1376.
8. Narayanan, A.; Shmatikov, V. Robust De-Anonymization of Large Sparse Datasets. In Proceedings of the 2008 IEEE Symposium on Security and Privacy (SP), Oakland, CA, USA, 18–21 May 2008; pp. 111–125.
9. Shokri, R.; Stronati, M.; Song, C.; et al. Membership Inference Attacks Against Machine Learning Models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–24 May 2017; pp. 3–18.
10. Albrecht, J.P. How the GDPR Will Change the World. *Eur. Data Prot. L. Rev.* **2016**, *2*, 287–289.
11. Voigt, P.; Von dem Bussche, A. *The EU General Data Protection Regulation (GDPR), A Practical Guide*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017.
12. U.S. Department of Health and Human Services *Summary of the HIPAA Privacy Rule*; U.S. Department of Health and Human Services: Washington, DC, USA, 2003.
13. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–487.
14. Erlingsson, Ú.; Pihur, V.; Korolova, A. Rappor: Randomized Aggregatable Privacy-Preserving Ordinal Response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014; pp. 1054–1067.
15. Kairouz, P.; McMahan, H.B.; Avent, B.; et al. Advances and Open Problems in Federated Learning. *Found. Trends Mach. Learn.* **2021**, *14*, 1–210.
16. Warner, S.L. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Am. Stat. Assoc.* **1965**, *60*, 63–69.
17. Mao, Y.; Ye, Q.; Wang, Q.; et al. Differential Privacy for Time Series: A Survey. *IEEE Data Eng. Bull.* **2024**, *47*, 67–92.
18. Wang, Z.; Yan, W.; Oates, T. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585.
19. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
20. Shazeer, N.; Mirhoseini, A.; Maziarz, K.; et al. Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. *arXiv* **2017**, arXiv:1701.06538.
21. Middlehurst, M.; Large, J.; Flynn, M.; et al. HIVE-COTE 2.0: A New Meta Ensemble for Time Series Classification. *Mach. Learn.* **2021**, *110*, 3211–3243.
22. Deng, H.; Runger, G.; Tuv, E.; et al. A Time Series Forest for Classification and Feature Extraction. *Inf. Sci.* **2013**, *239*, 142–153.
23. Ye, L.; Keogh, E. Time Series Shapelets: A New Primitive for Data Mining. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 947–956.
24. Arachchige, P.C.M.; Bertok, P.; Khalil, I.; et al. Local Differential Privacy for Deep Learning. *IEEE Internet Things J.* **2019**, *7*, 5827–5842.
25. Ye, Q.; Hu, H.; Li, N.; et al. Beyond Value Perturbation: Local Differential Privacy in the Temporal Setting. In Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications, Vancouver, BC, Canada, 10–13 May 2021; pp. 1–10.
26. Zhang, Y.; Ye, Q.; Chen, R.; et al. Trajectory Data Collection with Local Differential Privacy. *arXiv* **2023**, arXiv:2307.09339.
27. Ye, Q.; Hu, H.; Huang, K.; et al. Stateful Switch: Optimized Time Series Release with Local Differential Privacy. In Proceedings of the IEEE INFOCOM 2023-IEEE Conference on Computer Communications, New York, NY, USA, 17–20 May 2023; pp. 1–10.
28. Fan, L.; Xiong, L. An Adaptive Approach to Real-Time Aggregate Monitoring with Differential Privacy. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 2094–2106.
29. Zhang, W.; Jiang, N.; Yang, S.; et al. Federated Transfer Learning for Remaining Useful Life Prediction in Prognostics with Data Privacy. *Meas. Sci. Technol.* **2025**, *36*, 076107.

30. El Ouadrhiri, A.; Abdelhadi, A. Differential Privacy for Deep and Federated Learning: A Survey. *IEEE Access* **2022**, *10*, 22359–22380.
31. Yang, X.; Wang, T.; Ren, X.; et al. Survey on Improving Data Utility in Differentially Private Sequential Data Publishing. *IEEE Trans. Big Data* **2017**, *7*, 729–749.
32. Mu, S.; Lin, S. A Comprehensive Survey of Mixture-of-Experts: Algorithms, Theory, and Applications. *arXiv* **2025**, arXiv:2503.07137.
33. Fedus, W.; Zoph, B.; Shazeer, N. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *J. Mach. Learn. Res.* **2022**, *23*, 1–39.
34. Riquelme, C.; Puigcerver, J.; Mustafa, B.; et al. Scaling Vision with Sparse Mixture of Experts. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8583–8595.
35. Du, N.; Huang, Y.; Dai, A.M.; et al. Glam: Efficient Scaling of Language Models with Mixture-of-Experts. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA; 17–23 July 2022; pp. 5547–5569.
36. Liu, Z.; Luo, Y.; Li, B.; et al. Learning Soft Sparse Shapes for Efficient Time-Series Classification. *arXiv* **2025**, arXiv:2505.06892.
37. Rastogi, V.; Nath, S. Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, Indianapolis, IN, USA, 6–10 June 2010; pp. 735–746.
38. Papernot, N.; Song, S.; Mironov, I.; et al. Scalable Private Learning with PATE. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
39. Abadi, M.; Chu, A.; Goodfellow, I.; et al. Deep Learning with Differential Privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, Austria, 24–28 October 2016; pp. 308–318.