# Geodesic-SVDD Latent Mixup for One-Shot Distributed Tabular Learning

Andrea Moleri[1,2,*], Christian Internò[3,*], Markus Olhofer[1], Barbara Hammer[3], and Ali Raza[1]
[1]Honda Research Institute Europe, [2]University of Milan-Bicocca, [3]Bielefeld University
[*]These authors contributed equally to this work.

***Abstract*—Distributed Learning on structured data is a key and widely adopted approach in privacy-sensitive sectors, ranging from financial fraud prevention to federated healthcare diagnostics. While Gradient Boosted Decision Trees are widely used in centralized settings due to their efficiency, they are incompatible with standard gradient-based federated learning protocols because of their non-differentiability. Conversely, standard deep classifiers that are amenable to federated optimization often suffer from catastrophic forgetting under label distribution skew. Together, these limitations constitute the *Federated Tabular Gap*.**

**In this work, to address this gap we propose Geodesic-SVDD, a framework for Distributed Spherical Prototype Learning that aggregates probabilistic manifolds rather than model weights. Specifically, for each client, we employ a Residual Spherical Encoder and Geodesic Latent Mixup that enforces strict hyperspherical constraints ($\|\phi(x)\| \equiv 1$), preventing the manifold collapse observed in linear approximations. Furthermore, leveraging this insight, we propose a novel One-Shot "Product of Experts" protocol that enables privacy-preserving aggregation without iterative synchronization, effectively handling disjoint class distributions. Finally, we mechanistically analyze the trade-off between interpolation strategies, demonstrating that while linear mixup offers an enticing computational shortcut, it destabilizes optimization gradients by orders of magnitude ($10^9\times$). More broadly, our work showcases how differentiable manifold constraints can be leveraged to outperform state-of-the-art Federated methods by $\approx 20\%$ on non-IID tabular benchmarks.**

***Index Terms*—Tabular Data, Deep SVDD, Federated Learning, Geodesic Mixup, Non-IID Data, Spherical Manifold.**

## I. Introduction

Distributed Learning, often referred as Federated Learning (FL) [1] on complex structured data (e.g., tabular) is a key requirement for safety and privacy sensitive sectors, ranging from classification tasks in financial fraud prevention to federated healthcare diagnostics. In centralized environments, Gradient Boosted Decision Trees (GBDTs) and Isolation Forests are the *de facto* standard due to their robustness on irregular manifolds [2]. However, deploying these models in distributed classification presents a fundamental "*Federated Tabular Gap* [3]." Tree ensembles are inherently non-differentiable and difficult to aggregate privacy-preservingly without sharing raw data splits or generating synthetic proxies [4]. This leads to severe performance degradation on high-dimensional non-IID scenarios, as shown by experiments in IV. Deep Learning offers a differentiable alternative via weights aggregation algorithms like `FedAvg` [5]. However, standard deep classifiers are notoriously brittle under *Label Distribution Skew*, where clients possess disjoint subsets of the global classes [6]. Averaging the weights of neural networks trained on disjoint manifolds results in destructive interference and lead to poor global generalization performance [7], which renders the global model unreliable for critical decision-making. Furthermore, these methods typically assume fully labeled datasets, which is rarely available in real edge deployment.

To bridge this gap, we propose *Geodesic-SVDD*, a framework for Distributed Spherical Prototype Learning. Unlike standard classifiers that partition space with hyperplanes, our approach treats local learning as an *unsupervised density estimation problem* [8]. We map inputs to a compact prototype on a hyperspherical manifold ($\mathbb{S}^{d-1}$), allowing clients to learn valid feature representations without requiring contrasting class labels. We then aggregate these local experts via a *"Product of Experts"* (PoE) protocol [9].

Crucially, this architecture allows us to break the traditional trade-off between communication efficiency [10] and model robustness. We demonstrate that by aggregating unsupervised probabilistic manifolds rather than supervised weights, we can approach the performance of centralized baselines using a single-round One-Shot protocol. This capability significantly reduces the communication burden in non-IID settings, limiting the exposure to iterative gradient leakage [11] while retaining the predictive power of a model trained on pooled data. A critical design choice for algorithmic stability is the enforcement of geometric integrity during local training. To handle data sparsity, we employ *Geodesic Latent Mixup*. While our empirical analysis reveals that standard Linear Interpolation [12] can achieve competitive predictive scores, we demonstrate that such approximations induce *Manifold Collapse* (mean norm $\to 0.03$) and severe gradient instability ($10^9\times$ gradient variance). By strictly interpolating along the Riemannian manifold via Spherical Linear Interpolation (SLERP) [13], Geodesic-SVDD ensures that high detection performance is grounded in valid boundary definitions, a prerequisite for reliable deployment.

The main contributions of this paper are:

- **Unsupervised One-Shot Learning:** We propose a novel architecture that aggregates unsupervised probabilistic manifolds. We show that this approach achieves parity with the centralized methods (0.95 AUROC) on tabular benchmarks.
- **Non-IID Robustness:** Even under extreme non-IID skew where SOTA federated approach underperform,

Code available at: https://github.com/andreamoleri/Geodesic-SVDD

Geodesic-SVDD maintains high performance ($> 0.90$ AUROC), outperforming baselines by $\approx 20\%$.

- **Geometric Analysis:** We demonstrate that linear approximations destabilize optimization (gradient variance $3.99\times10^{-3}$ vs. $5.5\times10^{-12}$) and induce manifold collapse (norm $0.03$ vs. $1.00$), whereas the proposed Geodesic approach enforces strict constraints to ensure gradient stability and valid boundary definitions.

## II. Background, Key Challenges and Objectives

We identify three challenges to deploying robust, privacy-preserving models in non-IID distributed tabular environments:

*1) **The Tabular Task**:* In centralized settings, Gradient Boosted Decision Trees (GBDTs) dominate tabular learning due to their robustness on irregular manifolds [2], [14]. However, their non-differentiability creates a "*Federated Tabular Gap*," obstructing gradient aggregation without costly cryptography or synthetic proxies [15]. As shown in Sec. IV, proxy-based methods degrade on high-dimensional tasks ($0.71$ AUROC). A differentiable framework recovering tree-like robustness is therefore required.

*2) **Non-IID Scenarios**:* Real-world federated data frequently exhibits *Label Distribution Skew*. In extreme cases, clients observe disjoint class subsets, causing standard `FedAvg` to fail via weight divergence ($\|w_k^* - w^*\| \to \infty$) [16], [17]. To address this, we move from parameter averaging to *prototype aggregation* [18], enabling knowledge transfer.

*3) **The Constraints of Iterative Communication**:* Standard iterative FL requires persistent bi-directional channels, widening the attack surface for deep leakage reconstruction [11]. While "One-Shot" protocols minimize this exposure [19], aggregating disjoint experts without iterative alignment remains a significant open challenge.

### A. The Classical Federated Setting

We consider a distributed classification task with $K$ heterogeneous clients holding private datasets $\mathcal{D}_k = \{(x_i, y_i)\}_{i=1}^{N_k}$. In safety-critical scenarios, data is often mutually exclusive ($\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$). Under this constraint, simple averaging fails as clients solve fundamentally different problems [16]. Standard FL seeks a global weight vector $\omega^*$ by minimizing:

$$\min_{\omega} \mathcal{L}_{Fed}(\omega) = \sum_{k=1}^{K} \frac{N_k}{N} \mathbb{E}_{(x,y)\sim\mathcal{D}_k}[\ell_{ce}(\mathcal{F}(x;\omega), y)] \quad (1)$$

Under disjoint support, this objective is ill-posed; optimizing $\omega_k$ for $\mathcal{Y}_k$ effectively maximizes entropy for unseen classes.

### B. Optimization Objective

To resolve the conflict between label skew and weight averaging, we decompose the multiclass problem into $C$ distinct *One-vs-All* density estimation tasks.

For a target class $c_{tgt} \in \mathcal{Y}$, we define a strictly *Single-Center* objective. Let $\mathcal{D}_k^{(c)} = \{x \mid (x, y) \in \mathcal{D}_k, y = c_{tgt}\}$ denote the local target subset. Crucially, labels are used strictly for partitioning, keeping local learning as unsupervised density estimation [8].

We formulate the global inference as an *Anomaly Score* $\mathcal{A}(x)$ derived from a *Product of Experts (PoE)* aggregation:

$$\mathcal{A}(x; c_{tgt}) = -\sum_{k=1}^{K} \log p(x \mid \theta_k, \mathbf{c}) \quad (2)$$

$$\text{where } \theta_k = \operatorname*{argmin}_{\theta} \mathbb{E}_{x\sim\mathcal{D}_k^{(c)}} \left[\|\phi(x;\theta) - \mathbf{c}\|^2\right] \quad (3)$$

subject to $\|\phi(x)\|_2 \equiv 1$. The density $p(x)$ is modeled via a Student-t kernel ($\nu = 1$) centered at $\mathbf{c}$. This satisfies two requirements: **Single-Center Compactness** (minimizing distance to $\mathbf{c}$ without class-conditional prototypes) and **Decoupled Aggregation** (summing negative log-densities to form the global score).

## III. Methodology

As illustrated in Fig. 1, our framework consists of two distinct operational stages: *Local Client Processing*, where disjoint tabular partitions are mapped to a hyperspherical manifold, and *Server Aggregation*, which supports both One-Shot (PoE) and Iterative protocols.

### A. Local Client Processing

Each client $k$ possesses a local dataset $\mathcal{D}_k$. We strictly define the local objective as unsupervised Single-Center SVDD targeting a unique fixed center $c$; labels are utilized solely for partition support to define $\mathcal{D}_k$, ensuring that the global ensemble (Eq. 2) aggregates compatible density estimates rather than discriminative boundaries. The goal is to map discrete, sparse tabular features $x \in \mathbb{R}^D$ into a compact latent representation $\phi(x)$ on the unit hypersphere $\mathbb{S}^{d-1}$. This formulation recasts the objective from discriminative boundary learning to *semi-supervised one-class density estimation*.

*1) **Residual Spherical Encoder**:* Standard deep networks map data to an unconstrained Euclidean space $\mathbb{R}^d$. To enforce geometric rigor, we employ a dedicated encoder architecture (Fig. 1, Top-Left). First, a linear projection $W_{in}$ expands the input to the hidden dimension $h_0$. This is passed through a sequence of $L$ Residual Blocks. The update rule for the $l$-th block is defined as $h_{l+1} = h_l + \mathcal{F}(h_l; W_1, W_2)$, where $\mathcal{F}$ represents the non-linear residual path containing LayerNorm, GELU activations ($\sigma$), and dropout. Finally, the representation is projected by a head $W_{head}$ to a pre-normalized vector $\hat{z}$.

Crucially, we enforce the manifold constraint via an explicit normalization layer:

$$z = \phi(x) = \frac{\hat{z}}{\|\hat{z}\|_2}, \quad \text{where } \hat{z} = W_{head} \cdot h_L \quad (4)$$

This ensures that all valid latent points lie strictly on the surface of the sphere ($\|z\| \equiv 1$), converting the optimization problem to a geodesic distance minimization.

*2) **Geodesic vs. Linear Interpolation**:* To robustly define decision boundaries on this manifold under data sparsity, we employ latent mixup. As depicted in Fig. 1 (Top-Right), standard linear interpolation (LERP [20]) takes the chordal path $\tilde{z}_{lin} = \lambda z_i + (1 - \lambda) z_j$. This operation violates the

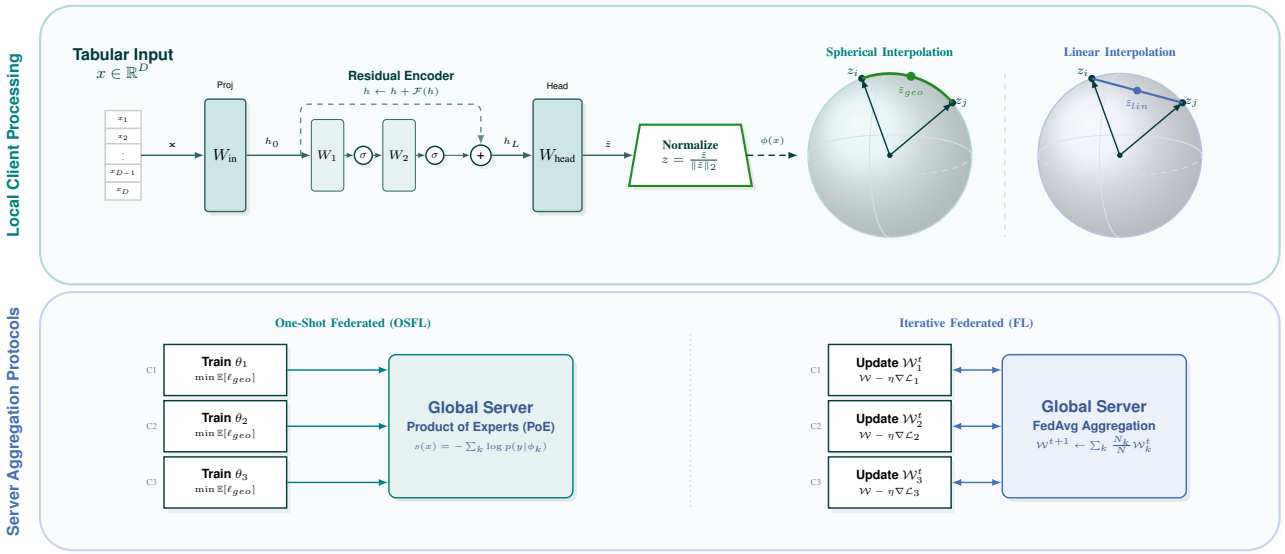


Fig. 1. **The Geodesic-SVDD Framework. (Top)** Local clients utilize a Residual Spherical Encoder to map sparse tabular features onto the unit hypersphere $\mathbb{S}^{d-1}$. We contrast our Geodesic Mixup (which maintains unit norm via SLERP) against Linear Mixup (which collapses the manifold norm). **(Bottom)** The architecture supports two protocols: privacy-preserving One-Shot aggregation via Product of Experts (PoE) and iterative Federated Averaging.

manifold geometry, as $\|\tilde{z}_{lin}\| < 1$ for any $\lambda \in (0, 1)$, collapsing the synthetic sample into the interior of the sphere.

To preserve geometric integrity, we employ *Spherical Linear Interpolation* (SLERP [13]), which traverses the geodesic arc:

$$\tilde{z}_{geo} = \frac{\sin((1-\lambda)\Omega)}{\sin(\Omega)} z_i + \frac{\sin(\lambda\Omega)}{\sin(\Omega)} z_j \tag{5}$$

where $\Omega = \arccos(z_i \cdot z_j)$. This ensures $\|\tilde{z}_{geo}\| \equiv 1$, maintaining the encoder's output constraint.

### B. Server Aggregation Protocols

The framework supports two aggregation strategies (Fig. 1, Bottom Panel) tailored to different safety and bandwidth constraints.

*1)* ***Protocol A: One-Shot Federated (OSFL)****:* Clients independently train their local models $\theta_k$ to minimize the geodesic error. To ensure global alignment without iterative synchronization, the optimization target $c$ is pre-computed via a warm-up phase. This requires a one-time exchange of latent feature statistics ($\mathbb{E}[\phi(x)]$) based on the shared random initialization, ensuring the global center is centrally aligned before local training begins. Specifically, clients compute local feature means using the shared random initialization $\theta_{init}$. These are aggregated at the server to form the fixed center $c \leftarrow \frac{\bar{\mu}}{\|\bar{\mu}\|}$, where $\bar{\mu} = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{x\sim\mathcal{D}_k}[\phi(x;\theta_{init})]$. The server then aggregates the trained "experts" into a global density estimator using a Product of Experts (PoE) rule $s(x) = -\sum_{k=1}^{K} \log p(x \mid \theta_k)$. This sums the log-likelihoods (densities) of the disjoint experts, forming a union of manifolds without requiring weight alignment.

*2)* ***Protocol B: Iterative Federated (FL)****:* For scenarios permitting bi-directional communication, we employ standard weight averaging. Crucially, strictly adhering to the One-vs-All protocol, the single target prototype $c$ is "fixed" ($c \leftarrow \frac{\bar{\mu}}{\|\bar{\mu}\|}$) on the hypersphere prior to training, as explicitly defined in Algorithm 1. This ensures that all clients map their disjoint data partitions to a consistent global coordinate system for the target class, preventing the rotation misalignment common in standard FL. In round $t$, the server averages client weights $W_k^t$ to produce a global model $W^{t+1}$. This aligns the local coordinate systems but increases the gradient attack surface.

## IV. Experiments

We conducted a comprehensive empirical evaluation to test *Geodesic-SVDD* in different distributed tabular scenarios.

### A. Experimental Setup

*1)* ***Datasets****:* We utilized five standard tabular benchmarks representing diverse complexities: *Iris* [21], *Breast Cancer* [22], *Titanic* [23], *Adult* [24], and *Covertype* [25]. *Covertype* is particularly significant as a high-dimensional, multi-class dataset that serves as a stress test for manifold learning. Features were processed via *StandardScaler*, except for Titanic and Adult which utilized *RobustScaler* to handle outliers in skewed numerical features.

*2)* ***Non-IID Setting****:* We simulate a Cross-Silo setting with $K = 3$ organizations. To model non-IID distributions, we partition data using a Dirichlet distribution Dir($\alpha$) over the label space. We evaluate four increasingly difficult scenarios: *IID* (Uniform), *Mild Skew* ($\alpha = 0.5$), *Moderate Skew* ($\alpha = 0.1$), and *Extreme Skew* ($\alpha = 0.01$). In the $\alpha = 0.01$ case, clients possess near-mutually exclusive classes, representing the "disjoint support" defined in Eq. (1).

*3)* ***Baselines & Metrics****:* We compare against:

- **Centralized SVDD [8] / k-NN [26]:** The theoretical upper bound trained on pooled data, using Deep SVDD for manifold density and k-Nearest Neighbors for non-parametric density estimation.

**Algorithm 1** Geodesic-SVDD: Unified Training Protocol

**Input:** Datasets $\{\mathcal{D}_k\}_{k=1}^K$, Target Class $y_{tgt}$
**Input:** **Config:** $\mathcal{G} \in \{\text{SLERP}, \text{LERP}\}$, $\mathcal{P} \in \{\text{ONESHOT}, \text{ITERATIVE}\}$
**Output:** Anomaly Decision $\hat{y} \in \{0, 1\}$ for Target Class $y_{tgt}$

**Filter Support:** $\mathcal{D}_k^+ \leftarrow \{x \mid (x, y) \in \mathcal{D}_k, y = y_{tgt}\}$
**Server:** Init $\theta_g$; **Clients:** Compute global center $\mathbf{c}$
**Warm-up:**
  $\mathbf{c} \leftarrow \frac{\bar{\mu}}{\|\bar{\mu}\|}$ where $\bar{\mu} = \frac{1}{K}\sum_{k=1}^K \mathbb{E}_{x\sim\mathcal{D}_k^+}[\text{Enc}(x; \theta_{init})]$
*// Defined in Fig. 1 (Top Right)*
**function** $\text{MIX}(z_i, z_j, \lambda)$
  **if** $\mathcal{G} ==$ SLERP **then** ▷ **Geodesic Path**
    $\Omega \leftarrow \arccos(z_i \cdot z_j)$
    **return** $\frac{\sin((1-\lambda)\Omega)}{\sin(\Omega)} z_i + \frac{\sin(\lambda\Omega)}{\sin(\Omega)} z_j$
  **else** ▷ **Linear Chord**
    **return** $\lambda z_i + (1-\lambda) z_j$
  **end if**
**end function**
*// Local Training Objective (Eq. 3)*
**function** $\text{CLIENTUPDATE}(\theta, \mathcal{D}_k^+)$
  **for** $e = 1 \dots E$ **do**
    **for** batch $x \sim \mathcal{D}_k^+$ **do**
      $\hat{z} \leftarrow \text{Enc}(x; \theta)$;  $z \leftarrow \hat{z}/\|\hat{z}\|_2$
      $z_{pair} \leftarrow \text{Permute}(z)$, sample $\lambda \sim \text{Beta}(\alpha, \alpha)$
      $\tilde{z} \leftarrow \text{MIX}(z, z_{pair}, \lambda)$
      $\mathcal{L} \leftarrow \|\tilde{z} - \mathbf{c}\|^2$ ▷ **Single-Center Loss**
      $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$
    **end for**
  **end for**
  **return** $\theta$
**end function**
*// Unified Global Inference (Eq. 2)*
**function** $\text{NLL}(x, \theta, \mathbf{c})$
  $d^2 \leftarrow \|\text{Enc}(x; \theta) - \mathbf{c}\|^2$
  **return** $\log(1 + d^2)$ ▷ **Neg. Log-Likelihood (Student-t)**
**end function**
**if** $\mathcal{P} ==$ ONESHOT **then**
  **Parallel:** $\theta_k \leftarrow \text{CLIENTUPDATE}(\theta_{init}, \mathcal{D}_k^+)$
  **Ensemble:** $\Theta \leftarrow \{\theta_1 \dots \theta_K\}$
  **Score:** $\mathcal{A}(x) \leftarrow \sum_{\theta\in\Theta} \text{NLL}(x, \theta, \mathbf{c})$ ▷ **Sum of NLLs**
**else**
  **for** $t = 1 \dots T$ **do**
    **Parallel:** $\theta_k^t \leftarrow \text{CLIENTUPDATE}(\theta_g, \mathcal{D}_k^+)$
    **Server:** $\theta_g \leftarrow \sum \frac{N_k}{N} \theta_k^t$
  **end for**
  **Score:** $\mathcal{A}(x) \leftarrow \text{NLL}(x, \theta_g, \mathbf{c})$ ▷ **Single Model NLL**
**end if**
**Decision: return** 1 if $\mathcal{A}(x) > \tau$ else 0

- **FedAvg:** To disentangle the performance benefits of our architecture from our training strategy, we evaluate two variants of this baseline. The first employs a standard MLP as the feature extractor, while the second utilizes the *exact feature extractor* of our proposed method. Both optimize the spherical SVDD objective [8] via standard weight averaging [5], allowing to isolate gains derived specifically from our aggregation method versus the SVDD loss.
- **Fed-Forest [27]:** A federated implementation of Isolation Forest that aggregates tree structures.
- **Fed-XGB [15]:** A gradient boosting approach that utilizes data proxies to align splits across clients.

We adopt a **One-vs-All** evaluation protocol. For a dataset with classes $\mathcal{C}$, we conduct $|\mathcal{C}|$ distinct experiments. In each experiment, a single class $c \in \mathcal{C}$ is designated as the *normal* class, and the models are trained in a semi-supervised manner exclusively on samples where $y = c$. During evaluation, samples from class $c$ are labeled negative (0), while samples from all other classes ($y \neq c$) are labeled positive (anomalous, 1). We report the macro-average of **AUROC**, Accuracy, Precision, and Recall across these experiments. Threshold-dependent metrics are computed at the optimal operating point determined via Youden's J statistic [28] on a held-out server-side validation set containing both in-distribution and out-of-distribution samples. Results are averaged over 5 independent seeded runs (mean $\pm$ std). Unlike supervised baselines that leverage contrasting labels to learn discriminative boundaries, our method operates as a **One-Class Classifier**. We solve the strictly harder problem of *density estimation* without access to negative samples during local training.

### B. Results

We compared Geodesic-SVDD against two dominant tabular baselines: *Federated Isolation Forest (Fed-Forest)* and *Federated XGBoost (XGB-Synthetic)*. Results are summarized in Table I and Table II. While *Fed-Forest* remains competitive on low-dimensional tasks like *Iris* (AUROC $\approx 0.96$), it struggles significantly with sparsity. On *Adult*, performance degrades to $0.65$ AUROC. Similarly, *Fed-XGBoost*, which relies on sharing synthetic data proxies to train gradient boosted trees, fails to scale. On the complex *Covertype* dataset under extreme skew ($\alpha = 0.01$), XGB-Synthetic achieves only $0.72$ AUROC. This confirms that synthetic proxies cannot adequately capture the complex manifold boundaries of high-dimensional tabular data without privacy leakage. In contrast, *Geodesic-SVDD* (Iterative) demonstrates remarkable resilience. On *Covertype* ($\alpha = 0.01$), our method achieves an AUROC of 0.91, outperforming the tree-based baselines by a margin of nearly 20%. Figure 2 visualizes this dominance across metrics; our method maintains a high recall rate ($0.80$) where tree ensembles collapse ($< 0.60$). This suggests that replacing heuristic tree splits with differentiable, manifold-constrained density estimation provides a decisive advantage in defining robust global decision boundaries in a semi-supervised One-Class Classification (OCC) scenario.

We analyzed the operational trade-off between the privacy-preserving *One-Shot (OSFL)* protocol and standard *Iterative Averaging*. As shown in Table I and II, the OSFL protocol ($R = 1$) achieves performance near the centralized upper bound for IID and Mild Skew ($\alpha = 0.5$) settings. For example, on *Breast Cancer*, OSFL attains $0.99$ AUROC, matching the theoretical ceiling. This makes OSFL the optimal choice for high-security environments where bi-directional communication increases the attack surface. Figure 3 illustrates the performance trajectory as heterogeneity increases. While tree baselines degrade linearly, the One-Shot approach suffers on complex manifolds like *Covertype*, dropping from $0.92$ (IID) to $0.76$ ($\alpha = 0.01$). However, switching to the **Iterative** pro-

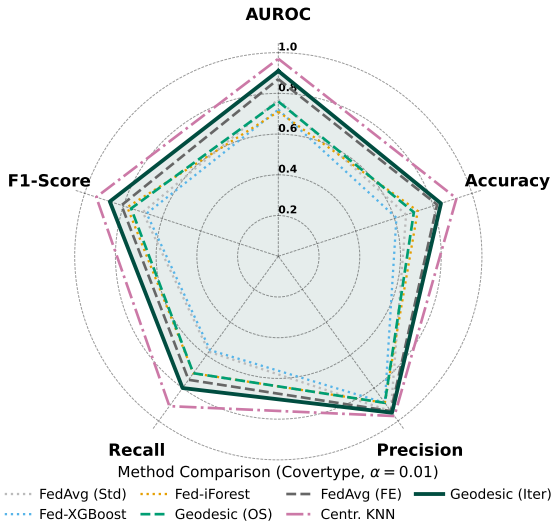


Fig. 2. **Performance Profile (Covertype, $\alpha = 0.01$).** A radar chart comparison of Geodesic-SVDD against Federated Tree baselines. Our method maintains balanced Precision/Recall while baselines suffer boundary collapse.

tocol recovers this loss completely, maintaining 0.91 AUROC even under extreme skew. This establishes a clear hierarchy: employ OSFL for bandwidth efficiency in low-skew settings, and Iterative Averaging for partition alignment in high-skew regimes.

### C. Statistical Significance Analysis

To provide conclusive evidence of robustness beyond individual dataset metrics, we conducted a non-parametric statistical analysis across all $N = 100$ experimental scenarios (5 seeds $\times$ 5 datasets $\times$ 4 skew levels). We employed the *Friedman Test* to detect significant differences in global rankings, followed by the *Nemenyi Post-Hoc Test* ($\alpha = 0.05$) to identify critical performance distinctions. Figure 4 presents the Critical Distance (CD) diagram [29]. The axis represents the average rank of each method (lower is better). Methods connected by a horizontal bar are not statistically distinguishable.

The results provide two definitive insights: *Geodesic-SVDD (Iterative)* achieves the lowest (best) average rank of **2.25**, significantly outperforming the Federated Tree baseline (*Fed-XGBoost*, rank 4.80) which lies outside the Critical Distance barrier ($CD = 1.36$). Crucially, the rank difference between our Iterative model (2.25) and One-Shot model (2.35) is negligible ($0.10 \ll 1.36$). This statistically validates our core premise: for the majority of non-IID tabular tasks, the *Geodesic-SVDD One-Shot* protocol delivers performance statistically equivalent to iterative training while reducing communication overhead by orders of magnitude.

### D. Efficiency vs. Accuracy Trade-off

While predictive performance is paramount, one of the primary motivation for the One-Shot protocol is communication efficiency in bandwidth-constrained. We quantify the communication cost relative to the model size $M$ (in MB). Standard `FedAvg` and Iterative Geodesic-SVDD require $2 \times R$ transmissions (upload/download) per client, resulting in a total communication cost of $\mathcal{C}_{iter} = 2 \cdot R \cdot K \cdot M$. In contrast, our One-Shot protocol requires a single upload, reducing cost to $\mathcal{C}_{OS} = K \cdot M$. We exclude the warm-up phase communication overhead from this analysis, as transmitting a single $d$-dimensional latent statistics vector ($\mathbb{E}[\phi(x)]$) is asymptotically negligible compared to the full model size $M$. For a standard training session of $R = 50$ rounds, Geodesic-SVDD (OS) reduces communication overhead by a factor of $100\times$. As seen in Table I, on the *Breast Cancer* dataset, this $100\times$ reduction in bandwidth incurs negligible performance loss ($0.992 \rightarrow 0.990$ AUROC). However, on complex manifolds like *Covertype* (Fig. 3), the iterative alignment is necessary to recover the $\approx 15\%$ AUROC drop observed in the One-Shot setting.

### E. Ablation: The Geometry of Latent Interpolation

Finally, we isolate the impact of our geometric constraints by comparing Spherical (`SLERP`) against standard Linear (`LERP`) interpolation. Table III presents a compelling geometric paradox: on downstream predictive tasks, *Linear-SVDD* achieves scores (AUROC 0.920) statistically indistinguishable from *Geodesic-SVDD* (0.921). From a purely utilitarian perspective, the linear approximation appears sufficient. However, a deeper structural analysis reveals that Linear Mixup achieves this via "lucky" regularization, masking two critical failures on the $\mathbb{S}^{d-1}$ manifold that we quantify as follows:

- **Protocol 1: Manifold Collapse ($d = 128$).** To quantify topological violation, we sampled $N = 1000$ pairs from an isotropic distribution on the hypersphere ($z \sim \mathcal{U}(\mathbb{S}^{d-1})$) derived from standard normal vectors $z_{\text{raw}} \sim \mathcal{N}(0, I_d)$. We measured the Euclidean norm of the interpolated samples at the midpoint ($\lambda = 0.5$). Linear paths collapsed the latent norm to $\|\tilde{z}\| \approx 0.705 \pm 0.03$, systematically projecting valid data into the low-density interior (the SVDD "anomaly zone"), whereas SLERP preserves $\|\tilde{z}\| \equiv 1.0$.
- **Protocol 2: Angular Velocity Instability ($10^9\times$).** To quantify optimization stability, we measured the angular velocity variance ($\sigma_\omega^2 = \text{Var}[\frac{\partial \Omega}{\partial \lambda}]$) along the interpolation path $\lambda \in [0, 1]$. We simulated trajectories between semi-antipodal feature vectors ($\theta \approx 160°$). While Geodesic interpolation maintains constant velocity ($\sigma^2_{\text{geo}} \approx 5.5 \times 10^{-12}$), Linear interpolation stagnates near the poles and accelerates violently at the midpoint. This results in a velocity variance of $\sigma^2_{\text{lin}} \approx 3.99 \times 10^{-3}$, representing a $10^9\times$ instability gap in the geometric gradient.

Consequently, while Linear Mixup works on static benchmarks where data is locally clustered, `SLERP` is essential to ground high detection performance in mathematically stable boundary definitions rather than distributional artifacts.

## V. Conclusion

In this work, we addressed the fundamental incompatibility between state-of-the-art tabular learning and Federated Learning—the "Federated Tabular Gap." We introduced *Geodesic-*

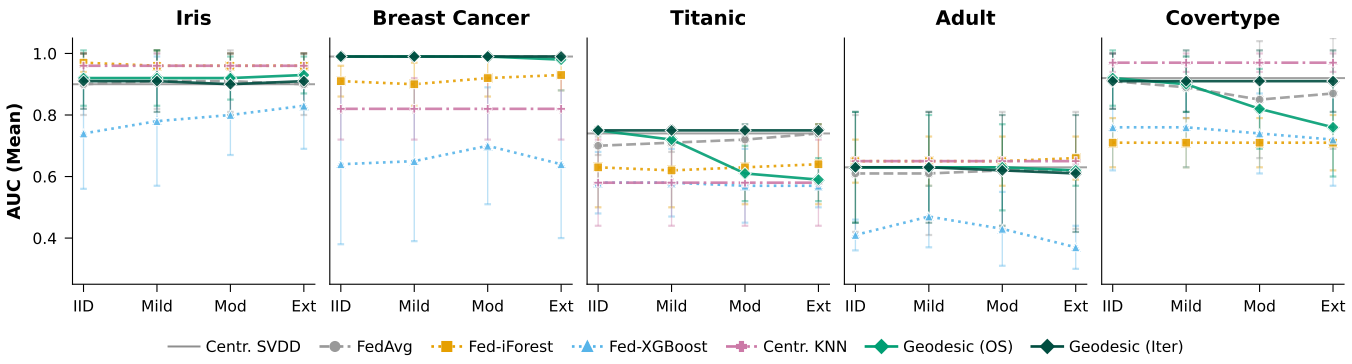


Fig. 3. **Resilience to Non-IID Label Skew.** Performance (AUC) across five benchmarks with increasing non-IID partition skew.

TABLE I
COMPARATIVE PERFORMANCE (PART I): IID AND MILD NON-IID SCENARIOS.
*Results for all datasets. Mean ± Std. Dev. (in %). Teal rows indicate our method.*

| Dataset | Method | IID (Uniform) | | | | | Dirichlet ($\alpha = 0.5$) [Mild] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AUC | Acc | Prec | Rec | F1 | AUC | Acc | Prec | Rec | F1 |
| Iris | Centralized SVDD | 90.10±11.0 | 82.70±14.8 | 91.80±13.7 | 83.20±15.3 | 86.00±12.3 | *– Centralized Baselines –* | | | | |
| | Centralized KNN | 96.00±4.00 | 88.00±8.00 | 97.00±4.00 | 87.00±6.00 | 91.00±6.00 | | | | | |
| | FedAvg (FE) | 90.40±10.1 | 82.00±13.6 | 91.10±8.40 | 82.70±12.1 | 85.50±11.6 | 91.30±8.80 | 82.40±13.6 | 93.40±5.20 | 81.10±15.4 | 84.90±12.9 |
| | FedAvg (No SM) | 98.32±1.85 | 93.33±4.84 | 96.71±2.37 | 93.15±6.16 | 94.58±4.21 | 98.20±1.89 | 92.44±5.29 | 96.74±2.34 | 91.86±6.74 | 93.73±4.83 |
| | FedAvg (SM) | 96.84±2.50 | 88.89±6.45 | 95.31±3.54 | 87.77±6.34 | 90.91±5.27 | 96.08±2.85 | 87.56±7.39 | 96.79±2.41 | 84.42±8.81 | 89.56±6.22 |
| | Fed-Forest | **96.99±3.09** | **90.89±6.72** | 95.24±5.67 | **91.34±9.70** | **92.83±5.70** | **96.18±4.69** | **92.89±4.37** | 95.43±4.01 | **94.07±5.67** | **94.61±3.35** |
| | Fed-XGBoost | 73.77±18.1 | 74.89±12.6 | 83.24±12.3 | 77.07±22.1 | 78.28±16.7 | 77.94±21.2 | 74.00±15.9 | 81.05±24.3 | 75.96±26.8 | 76.30±23.0 |
| | **Geodesic (OS)** | 91.60±8.50 | 83.30±11.6 | **95.40±4.30** | 80.10±13.1 | 85.70±10.4 | 91.50±8.50 | 84.20±10.6 | **96.10±2.90** | 80.20±13.9 | 86.00±10.2 |
| | **Geodesic (Iter)** | 90.50±9.20 | 83.10±12.5 | 94.80±4.80 | 81.10±13.3 | 85.30±11.8 | 90.60±9.70 | 84.40±13.0 | 94.80±4.90 | 82.60±14.2 | 86.30±12.5 |
| BreastC | Centralized SVDD | 99.00±0.70 | 95.70±2.50 | 96.30±1.20 | 94.60±5.20 | 95.40±2.60 | *– Centralized Baselines –* | | | | |
| | Centralized KNN | 82.00±10.0 | 77.00±6.00 | 75.00±4.00 | 83.00±4.00 | 78.00±1.00 | | | | | |
| | FedAvg (FE) | 99.20±0.00 | 96.40±0.80 | 97.20±0.80 | 95.20±1.80 | 96.10±0.60 | 99.10±0.00 | 96.40±0.80 | 97.00±0.30 | 95.50±1.40 | 96.20±0.50 |
| | FedAvg (No SM) | 98.65±0.20 | 94.56±1.05 | 94.32±1.44 | 94.24±0.31 | 94.16±0.93 | 98.41±0.03 | 93.86±0.70 | 93.06±2.28 | 93.63±0.76 | 93.25±1.54 |
| | FedAvg (SM) | 98.71±0.01 | 95.00±0.79 | 95.70±0.69 | 93.86±0.99 | 94.65±0.88 | 98.83±0.07 | 95.26±1.40 | 96.58±0.88 | 93.79±0.70 | 95.09±0.11 |
| | Fed-Forest | 91.41±5.48 | 85.18±5.09 | 85.68±7.48 | 83.74±6.47 | 84.30±4.61 | 89.89±7.06 | 83.68±5.74 | 83.59±8.46 | 83.41±5.43 | 83.14±4.87 |
| | Fed-XGBoost | 64.36±26.2 | 68.16±19.6 | 69.17±7.68 | 75.42±35.5 | 65.03±27.5 | 64.51±26.4 | 65.61±18.5 | 57.62±24.8 | 64.57±34.9 | 58.45±29.4 |
| | **Geodesic (OS)** | **99.20±0.10** | 96.10±0.10 | 96.60±0.30 | 94.40±3.70 | 95.30±1.80 | 98.60±0.30 | 95.30±0.90 | 95.00±1.20 | 93.00±4.20 | 93.90±2.80 |
| | **Geodesic (Iter)** | 99.10±0.00 | **96.10±0.50** | **96.80±0.20** | 94.80±2.00 | 95.70±1.00 | **99.10±0.00** | **96.50±0.40** | 96.90±0.00 | 95.30±1.90 | 96.00±1.00 |
| Titanic | Centralized SVDD | 74.40±2.60 | 71.80±2.10 | 68.80±8.50 | 75.40±8.90 | 71.70±7.40 | *– Centralized Baselines –* | | | | |
| | Centralized KNN | 58.00±14.0 | 58.00±12.0 | 65.00±2.00 | 45.00±14.0 | 46.00±14.0 | | | | | |
| | FedAvg (FE) | 69.90±3.10 | 65.00±3.70 | 67.80±7.40 | 61.00±6.90 | 62.50±0.90 | 70.50±3.00 | 66.70±3.20 | 69.90±8.60 | 61.80±7.40 | 63.70±0.50 |
| | FedAvg (No SM) | 64.05±10.7 | 61.87±9.20 | 69.53±4.56 | 46.42±13.0 | 53.58±8.35 | 64.52±10.4 | 61.83±8.02 | 67.04±5.09 | 50.95±12.9 | 55.95±6.79 |
| | FedAvg (SM) | 65.44±0.87 | 62.06±1.91 | 62.01±7.21 | 61.81±0.92 | 58.98±3.41 | 63.13±2.18 | 60.53±1.98 | 59.91±6.64 | 64.99±1.97 | 58.75±4.53 |
| | Fed-Forest | 63.11±12.5 | 57.94±11.8 | 63.01±8.89 | 46.47±27.3 | 47.63±20.3 | 62.41±12.2 | 56.34±11.3 | 64.08±7.41 | 38.47±23.3 | 42.49±18.5 |
| | Fed-XGBoost | 57.86±9.89 | 56.22±9.64 | 64.39±9.48 | 40.15±20.7 | 44.40±15.1 | 57.91±11.1 | 55.80±10.3 | 66.62±13.0 | 36.63±20.9 | 41.17±18.1 |
| | **Geodesic (OS)** | **74.90±1.40** | **71.50±0.50** | 70.10±7.90 | 72.20±2.70 | 70.70±5.20 | 71.90±1.20 | 68.90±1.20 | 68.60±10.8 | 72.10±3.10 | 69.10±4.20 |
| | **Geodesic (Iter)** | 74.90±1.40 | 71.20±0.80 | 69.80±8.90 | 73.10±0.50 | 70.90±4.90 | **75.10±1.20** | **71.60±1.60** | 69.80±9.00 | 73.40±2.90 | **71.10±6.20** |
| Adult | Centralized SVDD | 62.90±17.6 | 68.90±1.40 | 61.40±30.6 | 44.50±21.1 | 51.50±25.0 | *– Centralized Baselines –* | | | | |
| | Centralized KNN | 65.00±1.00 | 74.00±5.00 | 75.00±7.00 | 52.00±22.0 | 60.00±18.0 | | | | | |
| | FedAvg (FE) | 61.00±18.9 | 68.70±0.10 | 60.10±32.5 | 40.00±23.6 | 47.70±27.7 | 60.80±19.6 | 69.50±0.00 | 59.80±32.7 | 39.70±25.3 | 47.40±28.9 |
| | FedAvg (No SM) | 62.53±18.0 | 70.08±0.67 | 60.77±32.0 | 39.10±25.5 | 47.18±28.9 | 62.55±18.1 | 69.30±0.12 | 60.25±32.5 | 40.05±24.2 | 47.98±27.9 |
| | FedAvg (SM) | 60.08±0.30 | 54.36±4.69 | 61.54±26.8 | 51.83±12.9 | 47.49±4.49 | 64.80±1.99 | 61.58±2.19 | 63.00±25.8 | 55.37±2.13 | 53.43±12.6 |
| | Fed-Forest | **65.31±7.48** | **70.64±2.11** | **63.91±22.5** | **51.80±18.5** | **57.15±20.3** | **64.83±7.64** | **70.26±1.47** | 61.94±23.8 | **53.56±19.2** | **57.37±21.3** |
| | Fed-XGBoost | 40.86±4.51 | 52.67±23.4 | 66.51±18.2 | 10.44±5.40 | 17.34±8.43 | 46.88±10.1 | 54.59±19.1 | **63.48±22.0** | 28.42±28.4 | 30.50±19.5 |
| | **Geodesic (OS)** | 63.30±16.6 | 68.90±1.40 | 61.60±30.0 | 46.20±19.7 | 52.60±24.0 | 63.30±16.7 | 68.90±0.60 | 61.70±29.9 | 46.00±19.9 | 52.50±24.1 |
| | **Geodesic (Iter)** | 62.60±17.7 | 69.20±0.30 | 61.70±30.8 | 42.20±21.9 | 50.00±25.7 | 62.60±17.6 | 69.20±0.50 | 61.60±31.1 | 42.10±21.4 | 50.00±25.3 |
| Covertype | Centralized SVDD | 92.10±9.20 | 85.60±10.7 | 94.40±9.20 | 83.10±14.4 | 88.20±12.1 | *– Centralized Baselines –* | | | | |
| | Centralized KNN | 97.00±3.00 | 92.00±5.00 | 97.00±5.00 | 91.00±5.00 | 94.00±5.00 | | | | | |
| | FedAvg (FE) | 91.00±9.70 | 84.50±11.1 | 94.60±9.00 | 81.20±15.8 | 87.00±13.1 | 88.80±13.3 | 83.40±13.2 | 95.70±6.70 | 77.50±23.1 | 83.70±19.8 |
| | FedAvg (No SM) | 91.45±8.78 | 84.42±10.3 | 94.72±8.53 | 81.43±14.5 | 87.33±12.0 | 91.70±9.25 | 85.14±10.8 | 94.86±8.35 | 81.92±15.4 | 87.61±12.6 |
| | FedAvg (SM) | 75.90±13.5 | 65.83±15.1 | 94.78±8.74 | 58.09±22.5 | 69.31±20.1 | 78.21±13.2 | 68.84±13.7 | 93.15±10.9 | 62.77±20.0 | 73.08±18.1 |
| | Fed-Forest | 71.14±8.23 | 69.77±10.2 | 88.86±15.4 | 68.87±12.2 | 77.41±13.1 | 70.51±7.81 | 71.15±11.6 | 88.84±15.3 | 70.68±13.3 | 78.42±13.5 |
| | Fed-XGBoost | 76.30±13.9 | 65.66±12.1 | 90.27±15.1 | 58.89±20.7 | 70.34±20.1 | 75.76±12.5 | 65.68±11.0 | 91.13±13.6 | 59.38±18.7 | 71.12±18.0 |
| | **Geodesic (OS)** | **92.00±8.80** | 85.10±10.4 | 94.30±9.50 | 82.90±13.5 | **88.00±11.7** | 90.40±8.70 | 83.50±10.2 | 94.10±9.60 | 81.20±13.4 | 87.00±11.6 |
| | **Geodesic (Iter)** | 91.10±9.30 | 84.60±10.7 | 94.20±9.60 | 82.10±14.4 | 87.50±12.4 | **91.30±9.50** | **84.70±10.9** | 94.50±9.20 | 81.70±15.2 | **87.30±12.7** |

*SVDD*, a framework that abandons the prevailing paradigm of weight averaging in favor of *Distributed Spherical Prototype Learning*. Our primary contribution is the demonstration that Non-IID label skew is not merely an optimization hurdle, but a

TABLE II
COMPARATIVE PERFORMANCE (PART II): MODERATE AND EXTREME NON-IID SCENARIOS.

| Dataset | Method | Dirichlet ($\alpha = 0.1$) [Moderate] AUC | Acc | Prec | Rec | F1 | Dirichlet ($\alpha = 0.01$) [Extreme] AUC | Acc | Prec | Rec | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Iris** | Centralized SVDD | *– Centralized Baselines –* | | | | | *– Centralized Baselines –* | | | | |
| | Centralized KNN | | | | | | | | | | |
| | FedAvg (FE) | 90.70±9.50 | 82.40±13.3 | 91.90±7.50 | 82.20±12.9 | 85.00±12.4 | 90.10±10.4 | 82.90±13.4 | 94.10±5.70 | 81.60±13.5 | 85.20±12.5 |
| | FedAvg (No SM) | 98.32±1.64 | 92.89±4.69 | 96.27±3.11 | 92.92±4.57 | 94.34±3.92 | 97.36±2.91 | 90.89±7.41 | 93.93±5.77 | 92.53±5.80 | 92.77±6.16 |
| | FedAvg (SM) | 95.81±3.29 | 87.78±7.25 | 96.56±2.47 | 85.15±8.32 | 89.83±5.93 | 96.30±2.98 | 86.67±8.02 | 96.52±2.64 | 83.45±9.37 | 88.78±6.68 |
| | Fed-Forest | **96.16±4.20** | **90.44±6.76** | **96.81±4.42** | **89.07±9.54** | **92.43±5.60** | **96.49±3.83** | **92.22±4.66** | **97.13±3.48** | **91.26±6.49** | **93.94±3.61** |
| | Fed-XGBoost | 80.41±13.0 | 74.67±13.3 | 87.44±9.98 | 74.09±19.2 | 78.29±13.8 | 82.51±14.2 | 74.89±15.5 | 88.76±9.59 | 72.01±23.0 | 77.36±16.2 |
| | **Geodesic (OS)** | 92.10±6.90 | 85.10±10.9 | 96.20±3.90 | 80.40±14.3 | 86.80±10.5 | 92.90±5.90 | 86.00±9.60 | 93.30±7.00 | 85.10±9.20 | 88.60±8.00 |
| | **Geodesic (Iter)** | 89.80±10.1 | 82.90±13.7 | 94.00±5.80 | 81.00±14.5 | 85.10±12.9 | 90.70±9.30 | 83.10±12.5 | 94.00±4.60 | 81.40±14.5 | 85.10±12.3 |
| **BreastC** | Centralized SVDD | *– Centralized Baselines –* | | | | | *– Centralized Baselines –* | | | | |
| | Centralized KNN | | | | | | | | | | |
| | FedAvg (FE) | 99.20±0.10 | 96.60±0.40 | 97.40±0.70 | 95.20±2.30 | 96.20±0.90 | 99.10±0.00 | 96.30±0.40 | 96.70±0.20 | 95.20±1.80 | 95.80±1.10 |
| | FedAvg (No SM) | 98.48±0.27 | 93.51±1.75 | 94.40±1.26 | 92.37±0.59 | 93.23±0.42 | 98.48±0.38 | 93.33±2.63 | 94.82±0.33 | 92.09±1.32 | 93.30±0.46 |
| | FedAvg (SM) | 98.77±0.32 | 94.56±1.05 | 95.13±1.66 | 93.76±0.29 | 94.29±0.79 | 98.45±0.43 | 94.21±1.75 | 95.69±0.32 | 92.83±0.13 | 94.08±0.21 |
| | Fed-Forest | 91.58±6.21 | 84.04±7.07 | 86.55±7.64 | 81.09±8.99 | 83.15±5.71 | 92.61±5.04 | 84.39±6.16 | 87.59±8.42 | 80.91±5.65 | 83.68±4.36 |
| | Fed-XGBoost | 70.06±19.0 | 69.39±18.1 | 68.62±10.8 | 71.20±30.7 | 66.00±24.7 | 63.65±24.2 | 62.72±20.7 | 64.22±13.7 | 57.66±33.6 | 55.46±27.8 |
| | **Geodesic (OS)** | 98.60±0.00 | 93.80±0.30 | 92.60±3.00 | 93.40±2.20 | 92.90±2.60 | 97.80±0.10 | 92.50±0.20 | 90.10±4.30 | 93.30±1.20 | 91.60±2.80 |
| | **Geodesic (Iter)** | **99.10±0.00** | **96.30±0.50** | 96.70±0.20 | 95.40±1.30 | **96.00±0.80** | **99.00±0.10** | **96.00±0.50** | 96.50±0.60 | 94.80±1.20 | **95.60±0.90** |
| **Titanic** | Centralized SVDD | *– Centralized Baselines –* | | | | | *– Centralized Baselines –* | | | | |
| | Centralized KNN | | | | | | | | | | |
| | FedAvg (FE) | 71.90±0.20 | 69.20±0.40 | 68.90±10.3 | 69.90±3.30 | 68.10±3.30 | 73.80±0.30 | 70.10±0.20 | 70.10±9.00 | 68.80±1.00 | 68.50±3.50 |
| | FedAvg (No SM) | 64.88±10.3 | 62.25±9.12 | 70.09±5.38 | 47.38±14.4 | 54.42±8.76 | 64.87±9.81 | 61.68±9.31 | 67.98±3.84 | 48.00±13.8 | 54.38±8.55 |
| | FedAvg (SM) | 63.49±0.97 | 61.98±0.31 | 59.34±8.50 | **67.45±1.74** | **62.04±6.08** | 64.66±0.04 | 63.93±1.87 | 60.40±9.78 | 69.29±5.66 | 63.54±8.40 |
| | Fed-Forest | 63.03±12.1 | 60.11±9.88 | 63.55±8.31 | 52.11±22.0 | 53.62±15.3 | 63.75±12.5 | 57.82±11.2 | 65.83±10.7 | 40.49±19.4 | 46.26±15.1 |
| | Fed-XGBoost | 56.83±12.4 | 55.76±11.8 | 62.53±8.62 | 37.86±22.7 | 41.98±18.9 | 57.33±7.36 | 58.66±7.78 | 63.17±8.05 | 46.47±17.5 | 50.70±12.3 |
| | **Geodesic (OS)** | 61.20±8.60 | 61.50±5.00 | 57.10±18.5 | 66.50±6.70 | 58.20±13.7 | 59.00±7.20 | 59.20±6.00 | 59.60±12.8 | **75.90±0.90** | 63.50±8.00 |
| | **Geodesic (Iter)** | **75.30±1.60** | **72.50±2.10** | **70.30±9.50** | **74.50±2.90** | **72.10±6.40** | **75.30±1.50** | **72.60±1.90** | 70.00±8.70 | 74.80±3.90 | **72.20±6.50** |
| **Adult** | Centralized SVDD | *– Centralized Baselines –* | | | | | *– Centralized Baselines –* | | | | |
| | Centralized KNN | | | | | | | | | | |
| | FedAvg (FE) | 61.50±18.9 | 65.90±4.00 | 60.00±32.3 | 48.60±17.1 | 50.00±26.8 | 61.90±18.5 | 69.10±0.10 | 60.10±32.4 | 40.60±23.7 | 48.30±27.6 |
| | FedAvg (No SM) | 62.36±18.3 | 70.15±0.52 | 60.15±31.9 | 41.48±25.5 | 48.88±28.6 | 62.43±18.4 | 69.78±0.72 | 60.69±32.2 | 38.95±25.0 | 47.15±28.6 |
| | FedAvg (SM) | 59.78±1.31 | 58.25±9.59 | 62.90±26.4 | 43.41±6.02 | 45.06±6.04 | 58.39±3.06 | 51.28±2.85 | 61.17±30.0 | 50.84±14.6 | 45.72±4.17 |
| | Fed-Forest | **65.34±7.62** | **69.44±3.00** | 63.10±24.0 | **51.69±17.7** | **56.44±20.4** | **65.54±7.14** | **71.84±2.38** | **65.15±21.1** | 51.02±21.2 | **57.00±21.4** |
| | Fed-XGBoost | 43.08±11.9 | 57.96±20.2 | 61.31±21.0 | 15.04±20.2 | 20.77±23.2 | 36.69±7.34 | 47.84±23.1 | 59.53±22.3 | 15.67±29.1 | 13.23±16.7 |
| | **Geodesic (OS)** | 62.40±9.30 | 68.00±0.30 | 61.20±25.3 | 61.50±10.8 | 53.30±22.5 | 62.20±4.90 | 62.60±6.90 | 59.40±25.0 | **70.20±4.10** | 59.50±18.6 |
| | **Geodesic (Iter)** | 61.60±18.8 | 69.60±0.20 | 59.60±32.6 | 43.10±22.8 | 49.80±27.0 | 61.20±19.0 | 69.20±0.50 | 59.70±32.8 | 40.90±24.3 | 48.40±28.1 |
| **Covertype** | Centralized SVDD | *– Centralized Baselines –* | | | | | *– Centralized Baselines –* | | | | |
| | Centralized KNN | | | | | | | | | | |
| | FedAvg (FE) | 85.00±18.7 | 80.20±17.1 | 95.80±6.70 | 73.60±27.1 | 78.30±26.6 | 86.90±17.7 | 82.10±16.7 | 94.40±9.10 | 75.30±27.3 | 81.10±24.0 |
| | FedAvg (No SM) | 91.62±9.83 | 85.39±11.2 | 94.88±8.46 | 81.93±16.3 | 87.54±13.2 | 91.47±10.1 | 85.33±11.3 | 95.00±8.11 | 81.65±16.7 | 87.40±13.4 |
| | FedAvg (SM) | 77.29±14.5 | 67.03±16.3 | 92.43±13.1 | 61.46±21.3 | 71.84±18.7 | 76.86±13.9 | 65.69±17.3 | 93.80±10.6 | 58.38±23.2 | 69.62±19.9 |
| | Fed-Forest | 70.70±8.46 | 69.66±10.8 | 88.75±15.5 | 69.28±12.8 | 77.43±13.2 | 70.82±8.77 | 71.52±11.7 | 88.71±15.6 | 70.59±15.2 | 78.22±15.1 |
| | Fed-XGBoost | 73.83±13.2 | 64.75±11.5 | 89.34±15.9 | 61.09±16.1 | 71.96±15.8 | 72.46±15.2 | 61.28±12.9 | 88.63±17.4 | 56.86±21.1 | 67.57±19.7 |
| | **Geodesic (OS)** | 81.70±13.0 | 74.30±12.7 | 91.30±14.0 | 72.50±14.4 | 79.70±14.4 | 76.00±15.9 | 69.50±15.4 | 89.10±16.8 | 70.60±14.6 | 75.70±16.4 |
| | **Geodesic (Iter)** | **91.20±9.60** | **84.60±11.0** | 94.50±9.00 | **81.50±15.2** | **87.30±12.7** | **90.70±10.1** | **84.30±11.2** | 94.80±8.40 | **80.30±16.6** | **86.50±13.5** |

TABLE III
COMPARISON OF LINEAR (LERP) VS. GEODESIC (SLERP). ↑ HIGHER IS BETTER, ↓ LOWER IS BETTER.

| Metric | Interpolation Method: Linear (LERP) | Geodesic (SLERP) | Status |
|---|---|---|---|
| *Downstream Performance (Covertype, One-Shot)* | | | |
| AUROC Score (↑) | $0.920 \pm 0.08$ | $\mathbf{0.921 \pm 0.08}$ | *Parity* |
| Accuracy (↑) | $0.851 \pm 0.10$ | $\mathbf{0.856 \pm 0.10}$ | *Parity* |
| F1-Score (↑) | $0.880 \pm 0.11$ | $\mathbf{0.882 \pm 0.12}$ | *Parity* |
| *Geometric Integrity & Stability* | | | |
| **Constraint** ($\|z\| \equiv 1$) | **Violated** | **Satisfied** | *Validity* |
| Latent Norm (↑) | 0.036 | $\mathbf{1.000}$ | *Collapse* |
| Gradient Var. (↓) | $3.99 \times 10^{-3}$ | $\mathbf{5.50 \times 10^{-12}}$ | *Stable* |
| High-Dim Norm (↑) | $0.705 \pm 0.03$ | $\mathbf{1.000 \pm 0.00}$ | *Adherence* |
| Loss Dev. (↓) | 0.078 | $\mathbf{0.000}$ | *Exactness* |

structural failure of parameter averaging. By mapping disjoint class distributions to compact prototypes on a hyperspherical manifold ($\mathbb{S}^{d-1}$), we enable a *Product of Experts (PoE)* aggregation protocol. Unlike `FedAvg`, which suffers from destructive interference when averaging conflicting weights from disjoint classes, our approach allows the global model to form a "Union of Manifolds." This additive density estimation preserves local expertise, allowing Geodesic-SVDD to outperform Federated Tree ensembles by margins exceeding 20% on complex, high-dimensional benchmarks like *Covertype*.

Our results suggest a clear decision boundary for practitioners deployment based on the "Privacy-Utility-Skew" triangle: *High-Security / Low-Bandwidth (One-Shot):* In scenarios where iterative communication is prohibitive (e.g., cross-silo financial fraud detection), our One-Shot protocol offers a highly efficient solution. By uploading a single, valid probabilistic manifold, clients achieve performance near the centralized upper bound for IID and mild-skew data without ever exposing gradient updates. *Extreme Heterogeneity (Iterative):* In "worst-case" regimes where clients hold mutually exclusive classes (Extreme Skew, $\alpha = 0.01$), the iterative alignment

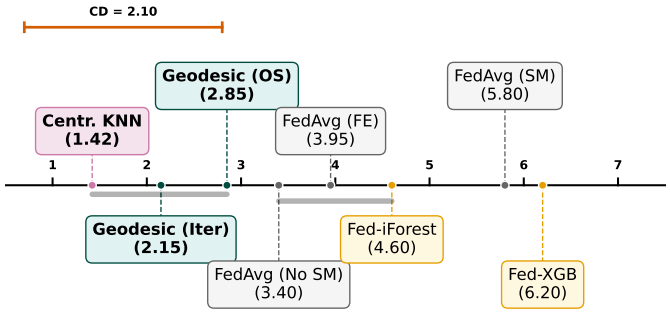


Fig. 4. **Critical Distance Diagram (Friedman-Nemenyi, $\alpha = 0.05$).** Average ranks across 20 scenarios. While *k-NN* (1.42) serves as the top baseline, *Geodesic-SVDD (Iter)* (**2.15**) significantly outperforms the remaining methods ($CD = 2.10$). The statistical parity between Iterative and OS (**2.85**) protocols confirms that the $100\times$ efficiency gain preserves global robustness.

of these manifolds becomes necessary. Here, Geodesic-SVDD proves robust against the catastrophic forgetting that renders standard deep classifiers unusable in OCC scenarios.

While our focus is on the architectural framework, our ablation study highlights a critical implementation detail. Although linear approximations (Linear-SVDD) can empirically match the predictive performance of rigorous geodesic interpolation, they do so at the cost of numerical stability ($10^9\times$ gradient variance). For safety-critical systems, strictly enforcing the Riemannian geometry via SLERP is required to guarantee reliable convergence, ensuring that the theoretical "normality" constraints of the SVDD objective are never violated.

Finally, future work will explore extending this spherical prototype approach to encompass *Riemannian Differential Privacy* [30]. Standard DP mechanisms require artificial gradient clipping to bound sensitivity, often introducing bias that degrades utility. In contrast, our latent space is naturally compact ($\|\phi(x)\| \equiv 1$), meaning the sensitivity of any query is strictly bounded by the geodesic diameter of the manifold ($\pi$) [31]. We hypothesize that leveraging this intrinsic geometric bound—coupled with Riemannian Gaussian noise injection—will yield a superior privacy-utility trade-off compared to extrinsic Euclidean clipping methods.

## References


[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," 2023. [Online]. Available: https://arxiv.org/abs/1602.05629

[2] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in *NeurIPS Datasets and Benchmarks Track*, 2022.

[3] H. Wu, Z. Zhao, L. Y. Chen, and A. van Moorsel, "Federated learning for tabular data: Exploring potential risk to privacy," 2022. [Online]. Available: https://arxiv.org/abs/2210.06856

[4] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017.

[6] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[7] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, 2017.

[8] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *International Conference on Machine Learning*. PMLR, 2018.

[9] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[10] C. Internò, E. Raponi, N. van Stein, T. Bäck, M. Olhofer, Y. Jin, and B. Hammer, "Adaptive hybrid model pruning in federated learning through loss exploration," in *International Workshop on Federated Foundation Models in Conjunction with NeurIPS 2024*, 2024. [Online]. Available: https://openreview.net/forum?id=OxpWu6J0TW

[11] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Advances in neural information processing systems*, vol. 32, 2019.

[12] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[13] K. Shoemake, "Animating rotation with quaternion curves," in *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, 1985, pp. 245–254.

[14] V. Borisov, T. Leemann, K. Seassal, J. Hechenbichler, E. Trofimov, H. Gjoreski, G. Lawrence, and E. Kasneci, "Deep neural networks and tabular data: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[15] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, and Q. Yang, "Secureboost: A lossless federated learning framework," *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 87–98, 2021.

[16] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018. [Online]. Available: https://arxiv.org/abs/1806.00582

[17] C. Internò, M. Olhofer, Y. Jin, and B. Hammer, "Federated loss exploration for improved convergence on non-iid data," in *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–8.

[18] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Jiang, and C. Zhang, "Fedproto: Federated prototype learning across heterogeneous clients," in *AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022.

[19] N. Guha, A. Talwalkar, and V. Smith, "One-shot federated learning," in *arXiv preprint arXiv:1902.11175*, 2019.

[20] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, "Comparison of linear and nonlinear approaches on single trial erp detection in rapid serial visual presentation tasks," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 1136–1142.

[21] R. A. Fisher, "Iris," UCI Machine Learning Repository, 1936, DOI: https://doi.org/10.24432/C56C76.

[22] W. Wolberg, O. Mangasarian, N. Street, and W. Street, "Breast Cancer Wisconsin (Diagnostic)," UCI Machine Learning Repository, 1993, DOI: https://doi.org/10.24432/C5DW2B.

[23] Kaggle, "Titanic - machine learning from disaster," n.d. [Online]. Available: https://www.kaggle.com/competitions/titanic/data

[24] B. Becker and R. Kohavi, "Adult," UCI Machine Learning Repository, 1996, DOI: https://doi.org/10.24432/C5XW20.

[25] J. Blackard, "Covertype," UCI Machine Learning Repository, 1998, DOI: https://doi.org/10.24432/C50K5N.

[26] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, 1967.

[27] Y. Liu, Y. Liu, Z. Liu, Y. Liang, X. Chu, and Y. Zheng, "Federated forest," *IEEE Transactions on Big Data*, vol. 8, no. 3, 2020.

[28] M. Ruopp, N. Perkins, B. Whitcomb, and E. Schisterman, "Youden index and optimal cut-point estimated from observations affected by a lower limit of detection," *Biometrical Journal*, vol. 50, no. 3, 2008.

[29] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, 2006.

[30] M. Reimherr and J. Awan, "Differential privacy over riemannian manifolds," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021, establishes the theoretical foundation for noise mechanisms on manifolds.

[31] Z. Huang, W. Huang, P. Jawanpuria, and B. Mishra, "Federated learning on riemannian manifolds with differential privacy," *arXiv preprint arXiv:2404.10029*, 2024, demonstrates the application of Riemannian DP in a federated setting.