DRIP-MPU: DISTILLED CONTEXT RETRIEVAL FOR FEW-SHOT TABULAR MPU CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot tabular multi-positive and unlabeled (MPU) classification requires distinguishing multiple positive classes and an unknown negative class using only a few labeled positives and an unlabeled pool. Scarce supervision limits task-specific PU/MPU learning, while the absence of labeled negatives prevents the direct use of tabular foundation models that require every class to be represented in context. We propose the framework DRIP-MPU, which formulates this task as learning to retrieve a multiclass context for a frozen TabPFN. TabPFN serves both as a teacher guiding context retrieval and as the final in-context classifier. Starting from labeled positives and geometrically selected diverse pseudo-negatives, a Feedback-Guided Context Retriever learns from confidence-weighted out-of-fold TabPFN predictions and geometric supervision. The retrieved pseudo-labeled examples update the context for the next round of teacher feedback, iteratively refining the context without updating TabPFN's parameters. Experiments on 25 tabular datasets show that DRIP-MPU achieves the highest average AUC and positive-class F1 among the evaluated PU/MPU baselines across 5-, 10-, and 15-shot settings. In the 5-shot setting, it improves upon the strongest baselines on these metrics by 7.7% and 46.1%, respectively. Code: `https://anonymous.4open.science/r/opal-forest-7c4d`.

1 INTRODUCTION

Tabular classification underpins applications in finance, healthcare, industrial systems, and business analytics. In many such applications, only a few instances are labeled for several classes of interest, while the remaining instances are unlabeled and may belong either to one of these target classes or to an unknown negative class. For example, in business-to-business prospect discovery, existing customers provide a few labeled examples for each target segment, whereas the remaining companies comprise both potential customers and companies outside all target segments. The task is to assign each unlabeled company to a target class or the unknown negative class. We refer to this problem as *few-shot tabular multi-positive and unlabeled (MPU) classification*.

Conventional tabular models, including gradient-boosted trees and deep neural networks, perform effectively only when all classes are represented in the labeled training data (Chen & Guestrin, 2016; Grinsztajn et al., 2022). Positive-Unlabeled (PU) (Chen et al., 2026; Wang et al., 2026) and Multi-Positive-Unlabeled (MPU) (Xu et al., 2017) methods instead derive classification supervision through risk estimation, self-training, distribution matching, or generative modeling. However, scarce positive examples provide weak support for estimating class priors and learning reliable decision boundaries. Methods relying on pseudo-labeled instances are also sensitive to the accuracy of their assigned labels (Zhao et al., 2026). In few-

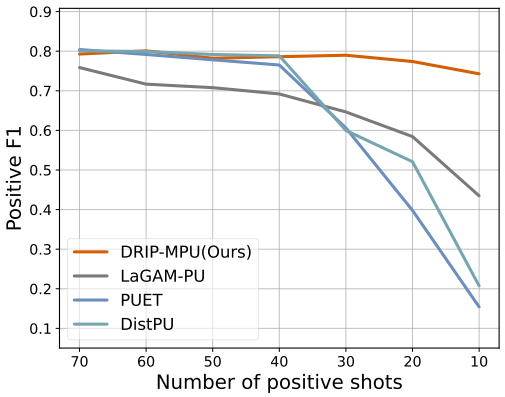


Figure 1: Positive F1 versus decreasing numbers of positive shots for PU learning methods on the Australian dataset.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

shot settings, these limitations result in a sharp decline in performance as the number of positive shots decreases, as illustrated in Figure 1.

Tabular foundation models offer a complementary source of transferable knowledge and perform strongly in low-data regimes (Van Breugel & Van der Schaar, 2024). For instance, TabPFN predicts query labels from labeled context examples without dataset-specific parameter updates (Hollmann et al., 2023; 2025; Grinsztajn et al., 2025; Ma et al., 2025). Its standard use, however, assumes that all classes are represented in the context. Few-shot MPU data violate this assumption because they contain no labeled negative examples and only a limited number of labeled examples for each positive class.

Consequently, few-shot tabular MPU classification falls outside the standard assumptions of conventional tabular models, existing PU/MPU methods, and tabular foundation models. We address this problem with the DRIP-MPU framework, short for *Distilled Context Retrieval and In-Context Prediction via Dual-Role TabPFN*. Unlike existing methods that directly train a final classifier, DRIP-MPU learns to construct a multiclass context for a pretrained tabular foundation model from scarce labeled positives and unlabeled data. Its idea is to use TabPFN both as a teacher that guides context construction and as the final in-context classifier. Starting from labeled positives and geometrically selected pseudo-negative examples, TabPFN provides out-of-fold predictions to train a Feedback-Guided Context Retriever (FGCR). The retriever constructs an updated pseudo-labeled context, which conditions the next round of teacher feedback. This iterative process adapts the context to the MPU task while keeping TabPFN frozen throughout.

Our main contributions are as follows:

- We introduce a learned context retrieval formulation of few-shot tabular MPU classification, transferring predictive knowledge from a frozen tabular foundation model to a retriever that constructs its multiclass context without labeled negatives.
- We develop a dual-role TabPFN framework that couples geometry-guided diverse negative initialization with feedback-guided context refinement. FGCR combines confidence-weighted out-of-fold distillation and geometric supervision to retrieve context examples for all classes.
- Experiments on 25 tabular datasets show that DRIP-MPU achieves the highest average AUC and positive-class F1 among the evaluated baselines across 5-, 10-, and 15-shot settings. In the 5-shot setting, it achieves relative improvements over the strongest baseline for each metric of 7.7% and 46.1%, respectively. With the same frozen TabPFN predictor, it also achieves a relative improvement of 19.2% over the strongest baseline in positive-class F1.

2 Related Work

Tabular Data Classification. Traditional tabular classifiers require dataset-specific fitting (Nelder & Wedderburn, 1972; Cover & Hart, 1967; Cortes & Vapnik, 1995; Chen & Guestrin, 2016; Prokhorenkova et al., 2018), while neural models such as TabNet and SAINT learn feature interactions end to end but retain the same requirement (Arik & Pfister, 2021; Somepalli et al., 2022). Tabular foundation models instead transfer knowledge acquired during pretraining. TabPFN performs in-context posterior-predictive inference without dataset-specific parameter updates (Hollmann et al., 2023; 2025). Subsequent models extend this paradigm through retrieval, learned contexts, real-table pretraining, and scalable in-context learning (Thomas et al., 2024; Feuer et al., 2024; Ma et al., 2025; Qu et al., 2025), with related PFN variants addressing forecasting and temporal drift (Dooley et al., 2023; Helli et al., 2024). LoCalPFN (Thomas et al., 2024) is related to our retrieval setting, but it retrieves from a fully labeled set and fine-tunes the classifier. We instead retrieve a pseudo-labeled multiclass context when the negative class has no labeled example. Despite their architectural differences, these methods assume access to labeled training or support data that covers every class. It cannot be met in our problem setting.

Positive-Unlabeled (PU) Learning. PU learning trains a binary classifier from labeled positives and an unlabeled positive–negative mixture (Elkan & Noto, 2008). The challenge is to recover reliable supervision from this mixture. One prominent line of work selects reliable examples using confidence, loss dynamics, predictive trends, or probability boundaries (Chen et al., 2020b; Luo et al., 2021; Zhu et al., 2023; Wang et al., 2023b; Li et al., 2024; Long et al., 2024). Another estimates objectives from all available PU data (Kiryo et al., 2017; Chen et al., 2020a; Garg et al., 2021;

Zhao et al., 2022; Jiang et al., 2023; Wilton et al., 2022). NcPU (Zhao et al., 2026) instead couples noisy-pair-robust representation alignment with iterative pseudo-target refinement. Subsequent variants address imbalance or covariate shift (Li et al., 2025a; Kumagai et al., 2025), relax labeling assumptions (Kato et al., 2019; Gong et al., 2022; Wang et al., 2023a), or introduce generative and adversarial models (Na et al., 2020; Papic et al., 2023; Hu et al., 2021).

MPU learning generalizes PU learning to several labeled positive classes and an unlabeled mixture that additionally contains an unknown negative class (Xu et al., 2017). Existing approaches develop multiclass risk estimators (Shu et al., 2020), identify unknown negatives through outlier detection, learn class prototypes, or target sparse prediction and minimum-risk shift (Li et al., 2025b). Extensions exploit graph structure or federated clients (Kim et al., 2025). These methods train a task-specific classifier or prototype model, and tabular variants generally assume substantially more labeled positives. Although these PU/MPU methods can perform classification without labeled negative examples, our experiments show that their performance deteriorates substantially under scarce positive supervision.

Tabular Few-Shot Learning. Tabular few-shot learning methods exploit unlabeled data through pseudo-task meta-learning, diffusion representations, or contrastive pretraining (Nam et al., 2023; Liu et al., 2024; Lee et al., 2025). LLM-based approaches transfer semantic knowledge through row serialization, generated features, cross-table pretraining, or feature-structure modeling (Hegselmann et al., 2023; Han et al., 2024; Gardner et al., 2024; Shi et al., 2025; Kang et al., 2026; Shi et al., 2026). These methods require labeled examples from every class. ScalePU (Dai et al., 2026) studies few-label binary PU classification but is only validated on images, while PUER (Wang et al., 2025) is specific to entity resolution. Thus, these methods do not address few-shot tabular MPU classification, where only a few labeled examples are available for each positive class and no labeled negative examples are provided.

3 Problem Setting

Few-shot tabular MPU classification. Let $\boldsymbol{x}_i \in \mathcal{X}$ denote a tabular instance with numerical and categorical features, and let $y_i \in \mathcal{Y}$ denote its class label. We apply a preprocessing function $\phi : \mathcal{X} \to \mathbb{R}^d$ that encodes categorical features and normalizes numerical features, yielding the vector representation $\phi(\boldsymbol{x}_i)$ for each instance $\boldsymbol{x}_i$. Following standard multi-positive and unlabeled (MPU) learning (Xu et al., 2017), $\mathcal{Y}$ is a set of k classes, consisting of $k-1$ positive classes denoted by $c_1, \ldots, c_{k-1}$ and one negative class denoted by c_{neg}, i.e., $\mathcal{Y} = \mathcal{Y}_p \cup \{c_{\text{neg}}\}$, $\mathcal{Y}_p = \{c_1, \ldots, c_{k-1}\}$, $|\mathcal{Y}| = k$. The negative class has no labeled examples. Each positive class $c \in \mathcal{Y}_p$ has n_c labeled examples, denoted by $\mathcal{P}_c = \{(\boldsymbol{x}_i, y_i = c)\}_{i=1}^{n_c}$, where each $\boldsymbol{x}_i$ is drawn from the class-conditional distribution $P(\boldsymbol{x} \mid y = c)$. The labeled data $\mathcal{P}$ therefore consists of the $k-1$ positive-class sets, with a total of n_p labeled positive examples, i.e., $\mathcal{P} = \bigcup_{c \in \mathcal{Y}_p} \mathcal{P}_c$, $n_p = |\mathcal{P}| = \sum_{c \in \mathcal{Y}_p} n_c$. The unlabeled pool contains n_u unlabeled instances, denoted by $\mathcal{U} = \{\boldsymbol{x}_i^u\}_{i=1}^{n_u} \sim \sum_{c \in \mathcal{Y}} \pi_c \cdot P(\boldsymbol{x} \mid y = c)$, where $\pi_c = P(y = c)$ is the class prior. The pool $\mathcal{U}$ contains instances from every class, but neither their labels nor the class priors are observed. Standard MPU learning has not been systematically studied under a fixed few-shot labeling budget. We consider its s-shot specialization by setting $n_c = s \ll n_u$ for every positive class $c \in \mathcal{Y}_p$. The *few-shot tabular MPU classification* task aims to predict the class label $y_i^u \in \mathcal{Y}$ for each unlabeled instance $\boldsymbol{x}_i^u \in \mathcal{U}$ using the training dataset $\mathcal{P} \cup \mathcal{U}$.

Learning objective and challenge. Rather than directly training a final classifier, we learn a context retriever r_θ. Given $(\mathcal{P}, \mathcal{U})$, it selects unlabeled instances and assigns each selected instance $\boldsymbol{x}_i^u$ a pseudo-label $\tilde{y}_i$, yielding $\mathcal{C}_\theta = r_\theta(\mathcal{P}, \mathcal{U}) \subseteq \{(\boldsymbol{x}_i^u, \tilde{y}_i) : \boldsymbol{x}_i^u \in \mathcal{U},\ \tilde{y}_i \in \mathcal{Y}\}$. For a query instance $\boldsymbol{x}_i^u$, we define its TabPFN context as $\mathcal{D}_{\theta,i} = \mathcal{P} \cup \{(\boldsymbol{x}_\ell^u, \tilde{y}_\ell) \in \mathcal{C}_\theta : \ell \neq i\}$. The query itself is excluded to prevent exposing its pseudo-label to TabPFN. Conditioned on $\mathcal{D}_{\theta,i}$, TabPFN produces the probability $P(y = c \mid \boldsymbol{x}_i^u, \mathcal{D}_{\theta,i})$ that $\boldsymbol{x}_i^u$ belongs to class c. The predicted label is $\hat{y}_i = \arg\max_{c \in \mathcal{Y}} P(y = c \mid \boldsymbol{x}_i^u, \mathcal{D}_{\theta,i})$.

An ideal retrieved context enables TabPFN to make accurate predictions on the unlabeled instances. We therefore seek a retriever whose constructed contexts minimize TabPFN's downstream multiclass prediction risk:

$$\theta^* \in \arg\min_\theta \mathcal{R}(\theta), \qquad \mathcal{R}(\theta) = \mathbb{E}\left[\frac{1}{n_u} \sum_{i=1}^{n_u} \mathcal{L}\big((P(y = c \mid \boldsymbol{x}_i^u, \mathcal{D}_{\theta,i}))_{c \in \mathcal{Y}}, y_i^u\big)\right], \tag{1}$$

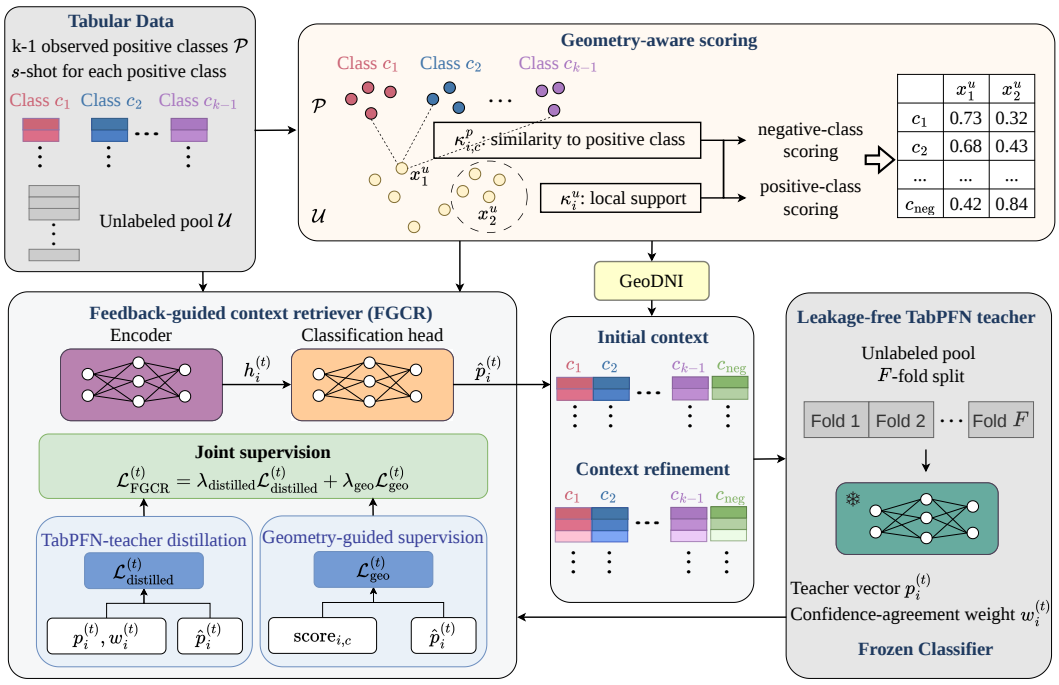


Figure 2: Architecture of the DRIP-MPU framework.

where $\mathcal{L}$ is a multiclass classification loss. Equation 1 characterizes an ideal retriever that selects and pseudo-labels instances such that the resulting context improves TabPFN's predictions on $\mathcal{U}$. However, this objective cannot be optimized directly because the true labels $\{y_i^u\}_{i=1}^{n_u}$ are unavailable. This challenge is particularly acute in the few-shot MPU setting, where labeled positives provide limited within-class coverage and no labeled negative examples are available. We address this challenge with the DRIP-MPU framework described in the next section.

4 DRIP-MPU Framework

4.1 Overview

To enable accurate classification in the few-shot tabular MPU setting, we propose the DRIP-MPU framework, short for *Distilled Context Retrieval and In-Context Prediction via Dual-Role TabPFN*. DRIP-MPU employs a frozen foundation model TabPFN in two complementary roles that form a self-reinforcing process. First, TabPFN serves as a teacher whose predictive knowledge is distilled into a context retriever that identifies informative pseudo-labeled examples from the unlabeled data. The trained retriever, in turn, constructs a more informative in-context set for TabPFN by retrieving pseudo-labeled examples. TabPFN then serves as the final in-context classifier, using this enriched context to predict the unlabeled instances. In this way, DRIP-MPU enables TabPFN to effectively help itself: its predictive knowledge guides context retrieval, while the improved retrieved context subsequently enhances its own in-context predictions. As illustrated in Figure 2, the framework consists of four components:

1. *Geometry-aware scoring and context initialization* computes a geometry-aware score for each unlabeled instance with respect to each class, uses these scores to select diverse pseudo-negative examples, and combines the selected negative examples with the labeled positive examples to construct the initial context $\mathcal{D}^{(0)}$ for TabPFN.
2. *Leakage-free TabPFN teacher* produces an out-of-fold class-probability vector for each unlabeled instance under the current context $\mathcal{D}^{(t)}$. These vectors provide soft supervision for training the context retriever through knowledge distillation and contribute to scoring candidates for the next-round context.
3. *Feedback-guided context retriever* (FGCR) is jointly trained under two complementary supervision signals: TabPFN-teacher distillation and geometry-guided supervision. For each unlabeled

instance, it outputs a class-probability vector, which is then used to select examples for each class and construct the context $\mathcal{D}^{(t+1)}$ for the next round. The second and third components alternate iteratively for T rounds, allowing the TabPFN teacher and the context retriever to progressively refine the retrieved context.

4. *Final prediction.* After T rounds, we obtain the final context $\mathcal{D}^{(T)}$. TabPFN predicts the label of each unlabeled instance conditioned on $\mathcal{D}^{(T)}$.

4.2 Geometry-Aware Scoring and Context Initialization

Geometry-aware scoring. We measure the similarity of an unlabeled instance $\boldsymbol{x}_i^u$ to positive class c by its maximum cosine similarity to the representations of the labeled examples in that class: $\kappa_{i,c}^p = \max_{(\boldsymbol{x}_j, y_j=c)\in\mathcal{P}_c} \cos(\phi(\boldsymbol{x}_i^u), \phi(\boldsymbol{x}_j))$. Motivated by nearest-neighbor density estimation (Loftsgaarden & Quesenberry, 1965), we measure its local support within the unlabeled pool using the average cosine similarity of its representation to those of its neighbors: $\kappa_i^u = \frac{1}{m}\sum_{j\in\mathcal{H}_m(i)} \cos\big(\phi(\boldsymbol{x}_i^u), \phi(\boldsymbol{x}_j^u)\big)$, where $\mathcal{H}_m(i)$ contains the m other unlabeled instances whose representations are most similar to $\phi(\boldsymbol{x}_i^u)$. A good negative candidate should exhibit low similarity to every observed positive class while receiving strong local support from the unlabeled pool, indicating that it is likely to belong to the negative class, rather than being noise or an outlier. For each unlabeled instance $\boldsymbol{x}_i^u$, we therefore define the *negative-class score* as:

$$\mathrm{score}_{i,c_{\mathrm{neg}}} = R\left[\left(1 - R\left(\max_{c\in\mathcal{Y}_p} \kappa_{i,c}^p\right)\right)\big((1-\lambda_\kappa) + \lambda_\kappa \cdot R(\kappa_i^u)\big)\right], \tag{2}$$

where $R(\cdot)$ rank-normalizes its argument across the unlabeled instances to $[0,1]$ (Conover & Iman, 1981) and $\lambda_\kappa \in [0,1]$ controls how strongly local support influences the negative-class score. A larger $\mathrm{score}_{i,c_{\mathrm{neg}}}$ indicates stronger evidence that $\boldsymbol{x}_i^u$ belongs to the negative class.

For each unlabeled instance $\boldsymbol{x}_i^u$, we define the *positive-class score* for class $c \in \mathcal{Y}_p$ as:

$$\begin{aligned} \mathrm{score}_{i,c} &= R\big[R(\kappa_{i,c}^p) \cdot \mathrm{sep}_{i,c} \cdot \big((1-\lambda_\kappa) + \lambda_\kappa \cdot R(\kappa_i^u)\big)\big], \\ \mathrm{sep}_{i,c} &= (1-\alpha) + \alpha \cdot R\left(\kappa_{i,c}^p - \max_{c'\in\mathcal{Y}_p\setminus\{c\}} \kappa_{i,c'}^p\right), \quad k>2, \quad c\in\mathcal{Y}_p. \end{aligned} \tag{3}$$

The class-similarity term $R(\kappa_{i,c}^p)$ is large when $\phi(\boldsymbol{x}_i^u)$ is similar to the labeled representations of positive class c. For $k > 2$, the class-separation score $\mathrm{sep}_{i,c}$ favors instances whose representations are more similar to class c than to every competing positive class. When $k = 2$, no competing positive class exists, so we set $\mathrm{sep}_{i,c} = 1$ and omit class separation. Setting $\alpha = 0$ removes the influence of class separation, whereas $\alpha = 1$ applies the rank-normalized class-separation score fully. The local-support term $(1-\lambda_\kappa) + \lambda_\kappa R(\kappa_i^u)$ has the same interpretation as in the negative-class score. Together, these terms favor instances whose representations are close to class c, well separated from other positive classes, and supported by neighboring unlabeled instances.

Context Initialization. Because no labeled negative examples are available, we infer the initial negative context from the unlabeled pool. However, selecting instances solely according to their negative-class scores may yield many instances from the same local region, resulting in a redundant and poorly representative set of negatives. We therefore propose *Geometry-guided Diverse Negative Initialization* (GeoDNI), a two-stage procedure that first restricts the candidate pool to geometrically plausible negative candidates and then promotes diversity among the candidates using maximal marginal relevance (MMR) (Carbonell & Goldstein, 1998).

We rank the unlabeled instances $\boldsymbol{x}_i^u \in \mathcal{U}$ by negative-class scores in descending order and retain the $\tau\%$ highest-scoring instances as the candidate pool $\mathcal{A}_{\mathrm{neg}} \subseteq \mathcal{U}$. Let $\gamma \le |\mathcal{A}_{\mathrm{neg}}|$ be the number of negative examples selected for initialization. We set $\mathcal{N}^{(0)} = \emptyset$ and run MMR for γ iterations. For each iteration $q > 1$ and each $\boldsymbol{x}_i^u \in \mathcal{A}_{\mathrm{neg}} \setminus \mathcal{N}^{(q-1)}$, we define its diversity score relative to the already selected instances as

$$\mathrm{div}^{(q-1)}(\boldsymbol{x}_i^u) = R\left[1 - \max_{\boldsymbol{x}_j^u\in\mathcal{N}^{(q-1)}} \cos\big(\phi(\boldsymbol{x}_i^u), \phi(\boldsymbol{x}_j^u)\big)\right], \qquad q>1. \tag{4}$$

At the first iteration, $q = 1$, no instance has been selected, so we set $\mathrm{div}^{(0)}(\boldsymbol{x}_i^u) = 0$. The MMR score of $\boldsymbol{x}_i^u$ is defined by combining its negative-class score with its diversity score:

$$\mathrm{MMR}^{(q)}(\boldsymbol{x}_i^u) = (1-\lambda_{\mathrm{div}}) \cdot R\big(\mathrm{score}_{i,c_{\mathrm{neg}}}\big) + \lambda_{\mathrm{div}} \cdot \mathrm{div}^{(q-1)}(\boldsymbol{x}_i^u), \qquad \lambda_{\mathrm{div}} \in [0,1), \tag{5}$$

where a larger λ_{div} places more weight on diversity. We then select the remaining candidate with the largest MMR score:

$$\boldsymbol{x}_{(q)}^{u} = \arg \max_{\boldsymbol{x}_i^u \in \mathcal{A}_{\text{neg}} \setminus \mathcal{N}^{(q-1)}} \text{MMR}^{(q)}(\boldsymbol{x}_i^u), \qquad \mathcal{N}^{(q)} = \mathcal{N}^{(q-1)} \cup \{\boldsymbol{x}_{(q)}^{u}\}. \tag{6}$$

After γ steps, let $\mathcal{N} = \mathcal{N}^{(\gamma)}$ denote the set of selected negative examples. We define the initial pseudo-labeled set by assigning these instances the negative label $\mathcal{C}^{(0)} = \{(\boldsymbol{x}_i^u, c_{\text{neg}}) : \boldsymbol{x}_i^u \in \mathcal{N}\}$, and combine it with the labeled positive set to form the initial context $\mathcal{D}^{(0)} = \mathcal{P} \cup \mathcal{C}^{(0)}$.

4.3 Leakage-Free TabPFN Teacher

The TabPFN teacher produces round-specific class-probability vectors for the unlabeled instances to supervise context retriever FGCR while TabPFN remains frozen. At round t, let $\mathcal{C}^{(t)}$ denote the pseudo-labeled set retrieved in the preceding round, and then the current context is $\mathcal{D}^{(t)} = \mathcal{P} \cup \mathcal{C}^{(t)}$. To satisfy the admissible-context requirement in Section 3, we generate the teacher vectors out of fold. Specifically, we distribute the instances in $\mathcal{C}^{(t)}$ evenly across $F \geq 2$ disjoint folds and randomly assign every remaining instance in $\mathcal{U}$ to these folds while keeping their sizes balanced, yielding $\mathcal{U} = \bigcup_{j=1}^{F} \mathcal{U}_j$. For a query $\boldsymbol{x}_i^u$ in fold $\mathcal{U}_j$, the corresponding context is $\mathcal{D}_j^{(t)} = \mathcal{P} \cup \left\{(\boldsymbol{x}_i^u, \tilde{y}_i^{(t)}) \in \mathcal{C}^{(t)} : \boldsymbol{x}_i^u \notin \mathcal{U}_j\right\}$. The context $\mathcal{D}_j^{(t)}$ excludes all pseudo-labeled instances in the query fold. TabPFN then produces the corresponding teacher vector as

$$\boldsymbol{p}_i^{(t)} = \left(P\left(y = c \mid \boldsymbol{x}_i^u, \mathcal{D}_j^{(t)}\right)\right)_{c \in \mathcal{Y}} \in \Delta^{k-1}, \qquad \boldsymbol{x}_i^u \in \mathcal{U}_j. \tag{7}$$

We refer to $\boldsymbol{p}_i^{(t)}$ as the *teacher vector* for $\boldsymbol{x}_i^u$ at round t. However, these teacher vectors are not equally trustworthy. We define a confidence-agreement weight for each teacher vector such that vectors with a larger top-two class-probability margin and stronger agreement with the geometry-aware scores receive larger weights. Let $p_i^{(t)}[1]$ and $p_i^{(t)}[2]$ denote the largest and second-largest entries of $\boldsymbol{p}_i^{(t)}$, respectively, and $p_{i,c}^{(t)}$ denote the probability of class c. The confidence-agreement weight is defined as:

$$w_i^{(t)} = \left(p_i^{(t)}[1] - p_i^{(t)}[2]\right) \left[\lambda_{\text{score}} + (1 - \lambda_{\text{score}}) \sum_{c \in \mathcal{Y}} p_{i,c}^{(t)} \cdot \text{score}_{i,c}\right], \tag{8}$$

where $\lambda_{\text{score}} \in [0, 1]$ controls the influence of geometry agreement. The teacher vectors $\{\boldsymbol{p}_i^{(t)}\}_{i=1}^{n_u}$ and their weights $\{w_i^{(t)}\}_{i=1}^{n_u}$ are used to supervise the FGCR at round t. Within each round t, FGCR is optimized for E epochs while keeping $\boldsymbol{p}^{(t)}$ and $w^{(t)}$ fixed.

4.4 Feedback-Guided Context Retriever

The Feedback-Guided Context Retriever (FGCR) maps any input instance $\boldsymbol{x}_i \in \mathcal{X}$ to a class-probability vector. Specifically, at round t, it first employs an encoder to obtain a latent vector $\boldsymbol{h}_i^{(t)} = f_\theta(\phi(\boldsymbol{x}_i)) \in \mathbb{R}^H$, where H denotes the dimension of the learned latent representation. A k-way linear classification head then projects $\boldsymbol{h}_i^{(t)}$ onto the k classes to obtain the class logits:

$$z^{(t)}(\phi(\boldsymbol{x}_i)) = \boldsymbol{W}\boldsymbol{h}_i^{(t)} + \boldsymbol{b}, \qquad \boldsymbol{W} \in \mathbb{R}^{k \times H}, \quad \boldsymbol{b} \in \mathbb{R}^k. \tag{9}$$

The logits are normalized using the softmax function to produce the class-probability vector:

$$\hat{\boldsymbol{p}}_i^{(t)} = \text{softmax}\left(z^{(t)}(\phi(\boldsymbol{x}_i))\right) = \left(\hat{p}_{i,c}^{(t)}\right)_{c \in \mathcal{Y}} \in \Delta^{k-1}, \tag{10}$$

where $\hat{p}_{i,c}^{(t)}$ is the probability that FGCR assigns to class $c \in \mathcal{Y}$ for input $\boldsymbol{x}_i$ at round t. FGCR is then trained using joint supervision consisting of two signals: TabPFN-teacher distillation and geometry-guided supervision.

TabPFN-Teacher Distillation. This distillation loss aligns the predicted class-probability vectors from FGCR with the teacher vectors from leakage-free TabPFN. Each teacher vector is weighted by

its confidence-agreement score.

$$\mathcal{L}_{\text{distilled}}^{(t)} = \frac{\sum_{i=1}^{n_u} w_i^{(t)} \, \mathrm{KL}\Big(\boldsymbol{p}_i^{(t)} \| \hat{\boldsymbol{p}}_i^{(t)}\Big)}{\sum_{i=1}^{n_u} w_i^{(t)} + \varepsilon}, \tag{11}$$

where $\mathrm{KL}(\cdot\|\cdot)$ is the Kullback–Leibler divergence and $\varepsilon > 0$ prevents division by zero. We set $\varepsilon = 10^{-8}$ in all experiments.

Geometry-Guided Supervision. For an unlabeled instance $\boldsymbol{x}_i^u \in \mathcal{U}$ and a class $c \in \mathcal{Y}$, we define the global geometric alignment loss by weighting the negative log-probability of each class according to the class score $\mathrm{score}_{i,c}$:

$$\mathcal{L}_{\text{ggeo}}^{(t)} = -\frac{1}{n_u} \sum_{i=1}^{n_u} \sum_{c \in \mathcal{Y}} \mathrm{score}_{i,c} \log \hat{p}_{i,c}^{(t)}. \tag{12}$$

For the pseudo-labeled instances $(\boldsymbol{x}_i^u, \tilde{y}_i^{(t)})$ in $\mathcal{C}^{(t)}$, we define the selected-context geometric alignment loss as:

$$\mathcal{L}_{\text{scgeo}}^{(t)} = -\frac{1}{|\mathcal{C}^{(t)}|} \sum_{(\boldsymbol{x}_i^u, \tilde{y}_i^{(t)}) \in \mathcal{C}^{(t)}} \sum_{c \in \mathcal{Y}} v_{i,c}^{(t)} \log \hat{p}_{i,c}^{(t)}. \tag{13}$$

where $v_{i,c}^{(t)}$ is the class-specific weight applied to $-\log \hat{p}_{i,c}^{(t)}$. For the pseudo-labeled class $c = \tilde{y}_i^{(t)}$, we set $v_{i,c}^{(t)} = \mathrm{score}_{i,\tilde{y}_i^{(t)}}$. For each alternative class $c \neq \tilde{y}_i^{(t)}$, we set $v_{i,c}^{(t)} = \big(1 - \mathrm{score}_{i,\tilde{y}_i^{(t)}}\big) \cdot \mathrm{score}_{i,c} / \sum_{c' \in \mathcal{Y} \setminus \{\tilde{y}_i^{(t)}\}} \mathrm{score}_{i,c'}$, which allocates the remaining weight among the alternative classes in proportion to their scores. Thus, a larger pseudo-label score places greater loss emphasis on the pseudo-labeled class, whereas a smaller score shifts more weight to alternative classes with larger scores.

Finally, the geometry-guided supervision combines the global geometric alignment and the selected-context geometric alignment:

$$\mathcal{L}_{\text{geo}}^{(t)} = \lambda_{\text{ggeo}} \mathcal{L}_{\text{ggeo}}^{(t)} + \lambda_{\text{scgeo}} \mathcal{L}_{\text{scgeo}}^{(t)}, \quad \lambda_{\text{ggeo}} + \lambda_{\text{scgeo}} = 1. \tag{14}$$

Joint Supervision. We integrate the two supervision signals into the joint FGCR objective:

$$\mathcal{L}_{\text{FGCR}}^{(t)} = \lambda_{\text{distilled}} \mathcal{L}_{\text{distilled}}^{(t)} + \lambda_{\text{geo}} \mathcal{L}_{\text{geo}}^{(t)}, \quad \lambda_{\text{distilled}} + \lambda_{\text{geo}} = 1. \tag{15}$$

Context Refinement. At round t, for each instance $\boldsymbol{x}_i^u \in \mathcal{U}$ and class $c \in \mathcal{Y}$, we define a refinement score that combines the class probability $\hat{p}_{i,c}^{(t)}$ predicted by FGCR, its margin over the strongest competing class, and the corresponding teacher probability $p_{i,c}^{(t)}$:

$$\mathrm{rescore}_{i,c}^{(t)} = \lambda_{\hat{p}} R\Big(\hat{p}_{i,c}^{(t)}\Big) + \lambda_{\text{diff}} R\left(\hat{p}_{i,c}^{(t)} - \max_{c' \in \mathcal{Y} \setminus \{c\}} \hat{p}_{i,c'}^{(t)}\right) + \lambda_p R\Big(p_{i,c}^{(t)}\Big), \quad \lambda_{\hat{p}} + \lambda_{\text{diff}} + \lambda_p = 1. \tag{16}$$

To construct the context for the next round, we reassign each unlabeled instance $\boldsymbol{x}_i^u \in \mathcal{U}$ a pseudo-label according to its largest round-t refinement score. For each class $c \in \mathcal{Y}$, we then retain the top ξ_c instances reassigned to c according to their refinement scores. This yields the round-specific retrieved set $\mathcal{C}^{(t+1)}$, from which the context for round $t+1$ is constructed as $\mathcal{D}^{(t+1)} = \mathcal{P} \cup \mathcal{C}^{(t+1)}$.

5 Experiments

5.1 Experimental Settings

Datasets. We conduct evaluations on 25 OpenML tabular datasets[1], comprising 15 binary-class datasets and 10 multi-class datasets with 4–10 classes. For each dataset, we perform a stratified $80\%/20\%$ train-test split, designate one class as the negative class, and treat the remaining classes

[1] `https://www.openml.org/search?type=data`.

as observed positive classes. The test set is used only for final evaluation. For each observed positive class, we randomly label 5, 10, or 15 training instances, while all remaining training instances constitute the unlabeled pool. Detailed dataset statistics are provided in Table 4 in Appendix A.

Baselines. We compare our method with 10 representative PU learning baselines applicable to tabular data, i.e., nnPU (Kiryo et al., 2017), UREA (Shu et al., 2020), AREA (Shu et al., 2020), BBE-PU (Garg et al., 2021), PUET (Wilton et al., 2022), Dist-PU (Zhao et al., 2022), CGenPU (Papic et al., 2023), PULDA (Jiang et al., 2023), LaGAM-PU (Long et al., 2024), and ScalePU (Dai et al., 2026). Except for UREA and AREA, the other eight baselines are designed specifically for binary classification and thus cannot be directly applied to multi-class datasets. Following the evaluation protocol of Shu et al. (2020), we extend these baselines to the multiclass setting by training a separate binary PU classifier for each of the $k-1$ observed positive classes. For each classifier, instances labeled as the corresponding positive class are treated as positive, while all other instances are treated as unlabeled. The final prediction is the positive class with the highest binary PU score, or the negative class if all $k-1$ PU scores are negative. We implement ScalePU, UREA, and AREA, as their source code are unavailable. For the other seven baselines, we use the implementations provided by PU-Bench (Chen et al., 2026). For baselines that require class-prior information (i.e., PUET, ScalePU, nnPU, Dist-PU, PULDA, UREA, and AREA), we employ KM2 (Ramaswamy et al., 2016) to estimate the positive-class priors from the data. For a comprehensive comparison, Appendix F additionally reports the results of these baselines using the true positive-class priors.

Metrics. For binary classification tasks, we use the positive-class F1 score and ROC-AUC as evaluation metrics. For multi-class classification tasks, we report positive-class Macro-F1 and macro one-versus-rest AUC (Macro-OvR AUC). For brevity, we denote the F1-based and AUC-based metrics as Pos.-F1 and AUC, respectively, in both settings. All experiments are repeated ten times with different random seeds for few-shot labeled-instance selection, and we report the mean and sample standard deviation across the ten runs.

5.2 Main Results

DRIP-MPU achieves the highest average performance across all shot settings. Table 1 summarizes results across 25 datasets, with complete per-dataset comparisons provided in Appendix F. The notation (KM2) following a method name indicates that its positive-class priors are estimated using KM2 (Ramaswamy et al., 2016). DRIP-MPU achieves the highest average AUC and Pos.-F1 across all three labeling budgets (5, 10, and 15 shots). In the most challenging 5-shot setting, it obtains 75.4% AUC and 56.7% Pos.-F1, outperforming the strongest baseline for each metric, PUET (KM2) and LaGAM-PU, by 5.4 and 17.9 percentage points, respectively. These gains correspond to relative improvements of 7.7% and 46.1%, respectively. These results demonstrate that DRIP-MPU substantially improves both ranking quality and positive-class prediction under scarce positive supervision.

DRIP-MPU achieves strong positive-class prediction with fewer labels. As the label budget increases from 5 to 15 shots, DRIP-MPU's AUC improves from 75.4% to 78.4%, while its Pos.-F1 increases from 56.7% to 61.9%. The baselines also benefit from additional labeled examples, narrowing the performance gap. In the 10-shot setting, DRIP-MPU outperforms the strongest baselines by 2.9 percentage points in AUC and 16.0 percentage points in Pos.-F1; in the 15-shot setting, these margins decrease to 2.3 and 14.1 percentage points, respectively. Notably, DRIP-MPU with only five shots achieves a higher average Pos.-F1 than every baseline with 15 shots, exceeding the best-performing baseline, LaGAM-PU (47.8%), by 8.9 percentage points. These results show that DRIP-MPU improves more over the competing methods when fewer positive labeled examples are available.

DRIP-MPU achieves superior ranking quality and positive-class prediction. AUC and Pos.-F1 capture two complementary aspects in classification. AUC measures how well the scores rank instances, while Pos.-F1 evaluates the positive-class predictions. Strong ranking alone can therefore coexist with a selected set that contains many false positives or misses many true positives. This distinction is evident in the 5-shot setting, where PUET (KM2) achieves a higher AUC than LaGAM-PU (70.0% vs. 67.9%) but a substantially lower Pos.-F1 (23.2% vs. 38.8%), with the same pattern observed at 10 and 15 shots. In contrast, DRIP-MPU achieves the highest average AUC and Pos.-F1 across all three settings.

Table 1: Average results across 25 datasets (mean ± SD over 10 seed-wise dataset averages). Bold and underlined values indicate the best and second-best results, respectively.

Method	5-shot		10-shot		15-shot	
	AUC (%)	Pos.-F1 (%)	AUC (%)	Pos.-F1 (%)	AUC (%)	Pos.-F1 (%)
DRIP-MPU (Ours)	**75.4** ± 1.1	**56.7** ± 0.9	**77.3** ± 1.3	**59.7** ± 1.0	**78.4** ± 1.2	**61.9** ± 0.5
BBE-PU	68.7 ± 1.2	2.4 ± 1.3	73.2 ± 1.3	11.2 ± 1.5	73.3 ± 1.7	20.6 ± 1.9
LaGAM-PU	67.9 ± 1.9	<u>38.8</u> ± 1.7	70.8 ± 2.1	<u>43.7</u> ± 1.8	73.0 ± 1.2	<u>47.8</u> ± 0.9
CGenPU	60.9 ± 2.0	32.4 ± 2.5	63.7 ± 2.0	39.2 ± 1.7	66.0 ± 1.2	42.8 ± 1.4
PUET (KM2)	<u>70.0</u> ± 1.2	23.2 ± 2.6	<u>74.4</u> ± 1.5	32.2 ± 4.2	<u>76.1</u> ± 1.5	38.4 ± 3.0
ScalePU (KM2)	66.2 ± 1.2	1.6 ± 0.7	70.5 ± 1.7	6.8 ± 1.6	72.7 ± 1.1	13.1 ± 2.1
nnPU (KM2)	60.5 ± 1.9	7.6 ± 2.7	64.9 ± 0.9	11.0 ± 4.5	66.1 ± 1.9	13.4 ± 3.1
Dist-PU (KM2)	68.4 ± 1.5	22.1 ± 2.7	72.9 ± 1.3	35.8 ± 3.6	75.3 ± 1.5	40.4 ± 2.2
PULDA (KM2)	68.3 ± 1.4	19.7 ± 2.5	72.1 ± 1.0	26.7 ± 3.2	74.3 ± 0.9	34.1 ± 3.1
UREA (KM2)	57.3 ± 0.9	22.3 ± 1.7	60.2 ± 1.7	29.1 ± 3.0	62.0 ± 1.0	33.3 ± 2.3
AREA (KM2)	59.7 ± 1.5	3.1 ± 0.9	63.5 ± 1.3	8.4 ± 1.4	65.7 ± 1.4	12.1 ± 0.8

5.3 Ablation Study

DRIP-MPU improves performance beyond using TabPFN as the final predictor. To determine whether DRIP-MPU's performance gains arise primarily from using TabPFN or from its context construction procedure, we compare different context construction methods under the same frozen TabPFN predictor. Specifically, we apply frozen TabPFN to the pseudo-labeled contexts generated by four representative baselines whose average 5-shot Pos.-F1 improves over their original results in Table 1. As shown in Table 2, DRIP-MPU achieves 56.75% Pos.-F1, outperforming the strongest variant, PUET (KM2) + TabPFN, by 9.15 percentage points. This margin demonstrates that DRIP-MPU's gains stem not only from using TabPFN for final prediction but also from its context construction procedure.

Ablation results validate the contributions of the four components. Table 3 evaluates the effects of GeoDNI, FGCR, $\mathcal{L}_{\text{distilled}}$, and $\mathcal{L}_{\text{geo}}$ on DRIP-MPU in the 5-shot setting. Removing GeoDNI reduces the average Pos.-F1 from 56.75% to 50.39%, a drop of 6.36 percentage points, while removing FGCR reduces it to 52.54%, a drop of 4.21 percentage points. These are the two largest decreases among the tested ablations, highlighting the importance of geometry-guided negative initialization and feedback-guided context retrieval. Removing $\mathcal{L}_{\text{distilled}}$ lowers Pos.-F1 by 2.05 percentage points to 54.70%, while removing $\mathcal{L}_{\text{geo}}$ lowers it by 1.16 percentage points to 55.59%. Thus, both supervision signals improve performance when the other is retained. Overall, the results support the joint use of negative initialization, feedback-guided retrieval, and the two supervision signals in the full framework.

Table 2: Baselines with frozen TabPFN.

Method	**Pos.-F1 (%)**
PUET (KM2) + TabPFN	47.60 ± 1.73
BBE-PU + TabPFN	44.93 ± 0.74
ScalePU (KM2) + TabPFN	44.27 ± 0.72
PULDA (KM2) + TabPFN	44.13 ± 1.60
DRIP-MPU (full)	**56.75 ± 0.90**

Table 3: DRIP-MPU component ablations.

Variant	**Pos.-F1 (%)**
w/o GeoDNI	50.39 ± 1.59
w/o FGCR	52.54 ± 1.52
w/o $\mathcal{L}_{\text{geo}}$	55.59 ± 0.99
w/o $\mathcal{L}_{\text{distilled}}$	54.70 ± 1.02
DRIP-MPU (full)	**56.75 ± 0.90**

6 Conclusion

We presented DRIP-MPU, a framework for few-shot tabular MPU classification that learns to construct multiclass contexts for a frozen TabPFN. By coupling TabPFN as both teacher and in-context classifier, DRIP-MPU uses teacher feedback to refine context retrieval and the retrieved context for subsequent feedback and final prediction. Geometry-guided negative initialization and feedback-guided retrieval enable this process using only scarce labeled positives and unlabeled data. Across 25 datasets, DRIP-MPU achieves the highest average AUC and positive-class F1 among the evaluated baselines under all labeling budgets.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation and editing code to implement the proposed method. We have not used generative AI tools to generate synthetic datasets; help develop theoretical models or conceptual frameworks; formulate mathematical claims; provide critical ingredients for proving mathematical claims; assist in the writing of proofs; propose or refine hypotheses; design or provide feedback on research methodology or experiments; clean or reformat datasets; support qualitative and thematic data analysis; or interpret results. Additionally, we used generative AI tools to draft parts of the research paper, summarize or analyse existing literature, brainstorm, search for information, edit the research paper to improve readability and format references. We have reviewed all AI-assisted work. All AI-assisted text was checked and revised by the authors, all literature-related information was checked against the original sources, and all AI-assisted code was manually reviewed and tested for consistency with the described method and reported experimental pipeline. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we release the complete source code through the anonymous GitHub repository (`https://anonymous.4open.science/r/opal-forest-7c4d`), which includes a detailed README with instructions for running the experiments. Table 4 lists all datasets used in our experiments and their corresponding OpenML identifiers. Dataset splits and parameter settings are detailed in Section 5.1 and Appendix B, respectively. Together, these materials provide the information needed to reproduce the reported experimental results.

REFERENCES

S. Ö. Arik and T. Pfister. TabNet: Attentive interpretable tabular learning. In *AAAI*, pp. 6679–6687, 2021.

Jaime G. Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *ACM SIGIR*, pp. 335–336, 1998.

H. Chen, F. Liu, Y. Wang, L. Zhao, and H. Wu. A variational approach for learning from positive and unlabeled data. In *NeurIPS*, pp. 14844–14854, 2020a.

Q. Chen, H. Zhang, L. Zhang, C. Li, J. Wang, and W. Wang. PU-Bench: A unified benchmark for rigorous and reproducible PU learning. In *ICLR*, 2026.

T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *ACM SIGKDD*, pp. 785–794, 2016.

X. Chen, W. Chen, T. Chen, Y. Yuan, C. Gong, K. Chen, and Z. Wang. Self-PU: Self boosted and calibrated positive-unlabeled training. In *ICML*, pp. 1510–1519, 2020b.

W. J. Conover and Ronald L. Iman. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3):124–129, 1981.

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

Y. Dai, X. Li, W. Wang, C. Li, G. Niu, and M. Sugiyama. Positive-unlabeled learning with extreme scarcity of labeled positives. In *ICML*, 2026.

S. Dooley, G. S. Khurana, C. Mohapatra, S. V. Naidu, and C. White. Forecastpfn: Synthetically-trained zero-shot forecasting. *NeurIPS*, pp. 2403–2426, 2023.

C. Elkan and K. Noto. Learning classifiers from only positive and unlabeled data. In *ACM SIGKDD*, pp. 213–220, 2008.

B. Feuer, R. T. Schirrmeister, V. Cherepanova, C. Hegde, F. Hutter, M. Goldblum, N. Cohen, and C. White. TuneTables: Context optimization for scalable prior-data fitted networks. In *NeurIPS*, pp. 83430–83464, 2024.

J. Gardner, J. C. Perdomo, and L. Schmidt. Large scale transfer learning for tabular data via language modeling. In *NeurIPS*, pp. 45155–45205, 2024.

S. Garg, Y. Wu, A. J. Smola, S. Balakrishnan, and Z. C. Lipton. Mixture proportion estimation and PU learning: A modern approach. In *NeurIPS*, pp. 8532–8544, 2021.

C. Gong, Q. Wang, T. Liu, B. Han, J. You, J. Yang, and D. Tao. Instance-dependent positive and unlabeled learning with labeling bias estimation. *IEEE TPAMI*, 44(8):4163–4177, 2022.

L. Grinsztajn, K. Flöge, O. Key, F. Birkel, P. Jund, B. Roof, B. Jäger, D. Safaric, S. Alessi, A. Hayler, M. Manium, R. Yu, F. Jablonski, S. B. Hoo, A. Garg, J. Robertson, M. Bühler, V. Moroshan, L. Purucker, C. Cornu, L. C. Wehrhahn, A. Bonetto, B. Schölkopf, S. Gambhir, N. Hollmann, and F. Hutter. TabPFN-2.5: Advancing the state of the art in tabular foundation models. *arXiv preprint arXiv:2511.08667*, 2025.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In *NeurIPS*, pp. 507–520, 2022.

S. Han, J. Yoon, S. O. Arik, and T. Pfister. Large language models can automatically engineer features for few-shot tabular learning. In *ICML*, pp. 17454–17479, 2024.

S. Hegselmann, A. Buendia, H. Lang, M. Agrawal, X. Jiang, and D. Sontag. TabLLM: Few-shot classification of tabular data with large language models. In *AISTATS*, pp. 5549–5581, 2023.

K. Helli, D. Schnurr, N. Hollmann, S. Müller, and F. Hutter. Drift-resilient tabpfn: In-context learning for temporal distribution shifts on tabular data. *NeurIPS*, pp. 98742–98781, 2024.

N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter. TabPFN: A transformer that solves small tabular classification problems in a second. In *ICLR*, 2023.

N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S. B. Hoo, R. T. Schirrmeister, and F. Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637 (8044):319–326, 2025.

W. Hu, R. Le, B. Liu, F. Ji, J. Ma, D. Zhao, and R. Yan. Predictive adversarial learning from positive and unlabeled data. In *AAAI*, pp. 7806–7814, 2021.

Y. Jiang, Q. Xu, Y. Zhao, Z. Yang, P. Wen, X. Cao, and Q. Huang. Positive-unlabeled learning with label distribution alignment. *IEEE TPAMI*, 45(12):15345–15363, 2023.

Inwon Kang, Parikshit Ram, Yi Zhou, Horst Samulowitz, and Oshani Seneviratne. Language model representations for efficient few-shot tabular classification. In *ACM Web Conference*, pp. 4069–4080, 2026.

M. Kato, T. Teshima, and J. Honda. Learning from positive and unlabeled data with a selection bias. In *ICLR*, 2019.

Junghun Kim, Hoyoung Yoon, Ka Hyun Park, and U Kang. Accurate graph-based multi-positive unlabeled learning via disentangled multi-view feature propagation. In *ACM SIGKDD*, pp. 1149–1159, 2025.

R. Kiryo, G. Niu, M. C. du Plessis, and M. Sugiyama. Positive-unlabeled learning with non-negative risk estimator. In *NeurIPS*, pp. 1675–1685, 2017.

A. Kumagai, T. Iwata, H. Takahashi, T. Nishiyama, K. Adachi, and Y. Fujiwara. Positive-unlabeled AUC maximization under covariate shift. In *ICML*, pp. 31876–31891, 2025.

Kyungeun Lee, Moonjung Eo, Hye-Seung Cho, Min-Kook Suh, Seoyoon Kim, Ye Seul Sim, Suhee Yoon, Sanghyu Yoon, and Woohyung Lim. Range-limited augmentation for few-shot learning in tabular data with comprehensive benchmark. In *ACM SIGKDD*, pp. 1262–1273, 2025.

C. Li, Y. Dai, L. Feng, X. Li, B. Wang, and J. Ouyang. Positive and unlabeled learning with controlled probability boundary fence. In *ICML*, pp. 27641–27652, 2024.

X. Li, Y. Dai, B. Wang, C. Li, J. Qu, and R. Guan. Balancing positive and negative classification error rates in positive-unlabeled learning. In *NeurIPS*, pp. 73290–73313, 2025a.

Zhongnian Li, Meng Wei, Peng Ying, and Xinzheng Xu. ESA: Example sieve approach for multi-positive and unlabeled learning. In *ACM WSDM*, pp. 439–447, 2025b.

Ruoxue Liu, Linjiajie Fang, Wenjia Wang, and Bing-Yi Jing. D2R2: Diffusion-based representation with random distance matching for tabular few-shot learning. In *NeurIPS*, pp. 36890–36913. Curran Associates, Inc., 2024.

D. O. Loftsgaarden and C. P. Quesenberry. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, 36(3):1049–1051, 1965.

L. Long, H. Wang, Z. Jiang, L. Feng, C. Yao, G. Chen, and J. Zhao. Positive-unlabeled learning by latent group-aware meta disambiguation. In *CVPR*, pp. 23138–23147, 2024.

C. Luo, P. Zhao, C. Chen, B. Qiao, C. Du, H. Zhang, W. Wu, S. Cai, B. He, S. Rajmohan, and Q. Lin. PULNS: Positive-unlabeled learning with effective negative sample selector. In *AAAI*, pp. 8784–8792, 2021.

J. Ma, V. Thomas, R. Hosseinzadeh, A. Labach, H. Kamkari, J. C. Cresswell, K. Golestan, G. Yu, A. L. Caterini, and M. Volkovs. TabDPT: Scaling tabular foundation models on real data. In *NeurIPS*, pp. 172692–172722, 2025.

B. Na, H. Kim, K. Song, W. Joo, Y.-Y. Kim, and I.-C. Moon. Deep generative positive-unlabeled learning under selection bias. In *ACM CIKM*, pp. 1155–1164, 2020.

Jaehyun Nam, Jihoon Tack, Kyungmin Lee, Hankook Lee, and Jinwoo Shin. STUNT: Few-shot tabular learning with self-generated tasks from unlabeled tables. In *ICLR*, 2023.

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *JRSS A*, 135(3):370–384, 1972.

A. Papic, I. Kononenko, and Z. Bosnic. Conditional generative positive and unlabeled learning. *ESWA*, 224:120046, 2023.

L. O. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. CatBoost: Unbiased boosting with categorical features. In *NeurIPS*, pp. 6639–6649, 2018.

Jingang Qu, David Holzmüller, Gaël Varoquaux, and Marine Le Morvan. TabICL: A tabular foundation model for in-context learning on large data. In *ICML*, pp. 50817–50847, 2025.

H. G. Ramaswamy, C. Scott, and A. Tewari. Mixture proportion estimation via kernel embeddings of distributions. In *ICML*, pp. 2052–2060, 2016.

R. Shi, H. Gu, H. Ye, Y. Dai, X. Shen, and X. Wang. Latte: Transferring LLMs' latent-level knowledge for few-shot tabular learning. In *IJCAI*, pp. 6173–6181, 2025.

Ruxue Shi, Yili Wang, Mengnan Du, Hangting Ye, Yi Chang, and Xin Wang. TAROT: Task-adaptive refinement of LLM-prior graphs for few-shot tabular learning. In *ACM SIGKDD*, pp. 4229–4239, 2026.

Senlin Shu, Zhuoyi Lin, Yan Yan, and Li Li. Learning from multi-class positive and unlabeled data. In *IEEE ICDM*, pp. 1256–1261, 2020.

G. Somepalli, A. Schwarzschild, M. Goldblum, C. B. Bruss, and T. Goldstein. SAINT: Improved neural networks for tabular data via row attention and contrastive pre-training. In *NeurIPS Workshop: TRL*, 2022.

V. Thomas, J. Ma, R. Hosseinzadeh, K. Golestan, G. Yu, M. Volkovs, and A. Caterini. Retrieval & fine-tuning for in-context tabular models. In *NeurIPS*, pp. 108439–108467, 2024.

Boris Van Breugel and Mihaela Van der Schaar. Position: Why tabular foundation models should be a research priority. In *ICML*, pp. 48976–48993, 2024.

W. Wang, D.-D. Wu, M. Li, J. Zhang, G. Niu, and M. Sugiyama. Accessible, realistic, and fair evaluation of positive-unlabeled learning algorithms. In *ICLR*, 2026.

X. Wang, H. Chen, T. Guo, and Y. Wang. PUe: Biased positive-unlabeled learning enhancement by causal inference. In *NeurIPS*, pp. 19783–19798, 2023a.

X. Wang, W. Wan, C. Geng, S. Li, and S. Chen. Beyond myopia: Learning from positive and unlabeled data through holistic predictive trends. In *NeurIPS*, pp. 67589–67602, 2023b.

Y. Wang, M. Yan, and W. Wang. PUER: Boosting few-shot positive-unlabeled entity resolution with reinforcement learning. In *EMNLP Findings*, pp. 24567–24579, 2025.

J. Wilton, A. M. Y. Koay, R. K. L. Ko, M. Xu, and N. Ye. Positive-unlabeled learning using random forests via recursive greedy risk minimization. In *NeurIPS*, pp. 24060–24071, 2022.

Yixing Xu, Chang Xu, Chao Xu, and Dacheng Tao. Multi-positive and unlabeled learning. In *IJCAI*, pp. 3182–3188, 2017.

H. Zhao, Z. Tu, Z. Zheng, W. Wang, J. Wang, R. Feagin, and W. Jiao. Noisy-Pair: Robust representation alignment for positive-unlabeled learning. In *ICLR*, 2026.

Y. Zhao, Q. Xu, Y. Jiang, P. Wen, and Q. Huang. Dist-PU: Positive-unlabeled learning from a label distribution perspective. In *CVPR*, pp. 14441–14450, 2022.

Z. Zhu, L. Wang, P. Zhao, C. Du, W. Zhang, H. Dong, B. Qiao, Q. Lin, S. Rajmohan, and D. Zhang. Robust positive-unlabeled learning via noise negative sample self-correction. In *ACM SIGKDD*, pp. 3663–3673, 2023.

A Detailed Description of Datasets

This section provides detailed information about the 25 OpenML classification datasets used in our experiments. As summarized in Table 4, the benchmark suite contains 15 binary-classification datasets and 10 multiclass-classification datasets. The datasets cover a broad range of problem sizes, with 208–9,873 instances, 4–81 features, and 2–10 classes. For each dataset, we report the label designated as the negative class in our experimental setting, and the corresponding OpenML dataset ID to facilitate unambiguous dataset retrieval and reproducibility.

Table 4: Statistics of the 25 OpenML classification datasets used in our experiments. The final column gives the OpenML dataset ID.

Dataset	#Feat.	#Inst.	#Classes	Negative label	OpenML ID
Binary classification					
Australian	14	690	2	0	40981
blood-transfusion	4	748	2	1	1464
climate-model-simulation-crashes	20	540	2	0	40994
credit-approval	15	690	2	-	29
credit-g	20	1,000	2	good	31
cylinder-bands	37	540	2	noband	6332
diabetes	8	768	2	tested_negative	37
dresses-sales	12	500	2	1	23381
kc2	21	522	2	no	1063
pc3	37	1,563	2	FALSE	1050
segment	19	2,310	2	N	958
sonar	60	208	2	Rock	40
spambase	57	4,601	2	0	44
tic-tac-toe	9	958	2	negative	50
wdbc	30	569	2	1	1510
Multiclass classification					
analcatdata_author	70	841	4	0	458
analcatdata_dmft	4	797	6	None	469
car	6	1,728	4	unacc	40975
GesturePhaseSegmentationProcessed	32	9,873	5	D	4538
first-order-theorem-proving	51	6,118	6	6	1475
mfeat-karhunen	64	2,000	10	0	16
mfeat-fourier	76	2,000	10	0	14
MiceProtein	81	1,080	8	c-CS-s	40966
satimage	36	6,430	6	1.	182
steel-plates-fault	27	1,941	7	Other_Faults	40982

B Implementation Details of DRIP-MPU

We use frozen TabPFN-2.5 for final inference. The hyperparameters are set to $m = 10$, $\lambda_\kappa = 0.60$, $\alpha = 0.65$, $\tau = 30\%$, $\gamma = \min\{32, \lceil 0.02|U| \rceil\}$ and $\lambda_{\text{div}} = 0.55$. Teacher probabilities are obtained using $F = 3$ pseudo-label-stratified out-of-fold folds, with $\lambda_{\text{score}} = 0.2$ for confidence-agreement weighting. We train FGCR for $E = 20$ epochs in each of $T = 30$ refinement rounds, resulting in 600 training epochs in total. We use $(\lambda_{\text{distilled}}, \lambda_{\text{geo}}) = (0.50, 0.50)$, with $(\lambda_{\text{ggeo}}, \lambda_{\text{scgeo}}) = (0.25, 0.75)$. For refinement, we use $(\lambda_{\hat{p}}, \lambda_p) = (0.625, 0.375)$ for binary datasets and $(\lambda_{\hat{p}}, \lambda_{\text{diff}}, \lambda_p) = (0.4375, 0.1875, 0.375)$ for multiclass datasets. For binary datasets, we set $\xi_c = \lfloor 0.05|U| \rfloor$ for the positive context, while $\xi_{c_{\text{neg}}}$ follows a $\lfloor 0.1|U| \rfloor \to \lfloor 0.15|U| \rfloor \to \lfloor 0.2|U| \rfloor$ progressive schedule over refinement rounds. For multiclass datasets, we set $\xi_c = 3n_c$ for five-shot settings and $\xi_c = 2n_c$ otherwise; for the negative class, we set $\xi_{c_{\text{neg}}} = \lfloor 0.15|U| \rfloor$. A complete parameter-sensitivity analysis is provided in Appendix E.

C Additional Ablation Studies

The gains are not explained by TabPFN alone. Table 5(a) reports the complete results of PU baselines combined with frozen TabPFN. Each PU baseline is trained using the labeled positive

Table 5: Complete frozen-TabPFN control and additional component ablations in the 5-shot setting. Bold and underlining in (a) indicate the best and second-best methods, respectively.

(a) Complete frozen-TabPFN control.

Method	AUC (%)	Pos.-F1 (%)
BBE-PU	69.3 ± 1.1	44.9 ± 0.7
LaGAM-PU	65.1 ± 1.8	40.8 ± 1.7
CGenPU	62.5 ± 2.4	37.4 ± 1.8
PUET (KM2)	$\underline{71.0} \pm 1.7$	$\underline{47.6} \pm 1.7$
ScalePU (KM2)	67.0 ± 1.2	44.3 ± 0.7
nnPU (KM2)	62.0 ± 1.6	37.0 ± 1.9
Dist-PU (KM2)	68.0 ± 1.6	44.1 ± 1.6
PULDA (KM2)	67.7 ± 1.7	44.1 ± 1.6
UREA (KM2)	64.4 ± 1.2	40.0 ± 1.0
AREA (KM2)	62.7 ± 1.2	38.2 ± 0.9
DRIP-MPU (Ours)	$\mathbf{75.4 \pm 1.1}$	$\mathbf{56.7 \pm 0.9}$

(b) Additional component ablations.

Variant	Pos.-F1 (%)
w/o GeoDNI	50.39 ± 1.59
w/o bootstrap MMR	56.37 ± 0.81
w/o FGCR	52.54 ± 1.52
w/o OOF	54.97 ± 1.37
w/o $\mathcal{L}_{\text{distilled}}$	54.70 ± 1.02
w/o $\mathcal{L}_{\text{geo}}$	55.59 ± 0.99
w/o $\mathcal{L}_{\text{ggeo}}$	56.19 ± 0.78
w/o $\mathcal{L}_{\text{scgeo}}$	55.70 ± 0.81
DRIP-MPU (full)	$\mathbf{56.75 \pm 0.90}$

examples and the unlabeled training pool. Its final trained model is retained as a scorer and applied to the real unlabeled training instances to produce pseudo labels and construct a pseudo-labeled context. This context is subsequently provided to frozen TabPFN-v2.5 for held-out test prediction.

This controlled setting holds the downstream prediction model fixed and therefore enables an objective evaluation of the advantage of our context construction procedure. Even when every method is equipped with the same strong frozen TabPFN predictor, DRIP-MPU retains a consistent aggregate advantage in both AUC and Pos.-F1. As shown in Table 5(a), DRIP-MPU achieves 56.75% Pos.-F1, outperforming the strongest variant, PUET (KM2) + TabPFN, by 9.15 percentage points. This shows that the performance gains do not arise merely from using TabPFN; rather, the complete DRIP-MPU framework constructs a higher-quality and more informative pseudo-labeled context for frozen TabPFN prediction.

Ablating GeoDNI and its MMR re-ranking. As shown in Table 5(b), removing GeoDNI reduces the average Pos.-F1 from 56.75% to 50.39%, a drop of 6.36 percentage points. This is the largest performance decrease among the component ablations, highlighting GeoDNI's critical role in constructing a reliable and informative initial negative context.

The MMR re-ranking in GeoDNI is designed to prevent the initial hidden-negative anchors from concentrating in the same local region, thereby improving their coverage of the negative distribution. As shown in Table 5(b), removing MMR produces almost no change in Pos.-F1. However, Macro-F1 decreases by 2.12 percentage points. This indicates that MMR does not primarily improve the recognition of observed positive classes. Instead, by constructing a more diverse set of initial negative anchors, it improves the representation and subsequent recognition of the hidden negative class, leading to more balanced multiclass predictions.

Effects of FGCR and leakage-free OOF mechanism. Removing FGCR reduces Pos.-F1 to 52.54%, a drop of 4.21 percentage points, confirming the importance of feedback-guided context refinement. We also replace leakage-free OOF teacher predictions with in-fold predictions, where the queried pseudo-labeled instance remains in the TabPFN context during teacher inference. This change decreases Pos.-F1 from 56.75% to 54.97%. The result supports the use of OOF feedback: excluding each queried instance from its own teacher context avoids self-conditioned predictions and prevents erroneous pseudo-labels from being repeatedly reinforced in teacher predictions.

Ablation of the four supervision losses. Removing $\mathcal{L}_{\text{distilled}}$ lowers Pos.-F1 by 2.05 percentage points to 54.70%, while removing $\mathcal{L}_{\text{geo}}$ lowers it by 1.16 percentage points to 55.59%. Thus, both supervision signals improve performance when the other is retained. We further ablate the two components of $\mathcal{L}_{\text{geo}}^{(t)}$: $\mathcal{L}_{\text{ggeo}}^{(t)}$ and $\mathcal{L}_{\text{scgeo}}^{(t)}$. The global term $\mathcal{L}_{\text{ggeo}}$ encourages retriever predictions over the unlabeled pool to align with the geometry-derived class support, whereas $\mathcal{L}_{\text{scgeo}}$ directly enforces geometric compatibility between selected pseudo-labeled examples and their assigned classes. Removing $\mathcal{L}_{\text{ggeo}}$ reduces Pos.-F1 from 56.75% to 56.19%, while removing $\mathcal{L}_{\text{scgeo}}$ leads to a larger drop to 55.70%. Thus, the global term regularizes the retriever over the unlabeled distribution, while the selected-context term more directly improves the context supplied to frozen TabPFN. Their combination achieves the best performance.

D PSEUDOCODE OF DRIP-MPU

The pseudocode of DRIP-MPU is summarized in Algorithm 1.

Algorithm 1: Pseudocode of DRIP-MPU

Input: labeled positive sets $\{\mathcal{P}_c\}_{c\in\mathcal{Y}_p}$, unlabeled set $\mathcal{U}$, $\mathcal{P}=\bigcup_{c\in\mathcal{Y}_p}\mathcal{P}_c$, $\mathcal{Y}=\mathcal{Y}_p\cup\{c_{\text{neg}}\}$, m, λ_κ, α, τ, λ_{div}, λ_{score}, $\lambda_{\text{distilled}}$, λ_{geo}, λ_{ggeo}, λ_{scgeo}, $\lambda_{\hat{p}}$, λ_{diff}, λ_p, $\{\xi_c\}_{c\in\mathcal{Y}}$, F, E, and T

Output: Final context $\mathcal{D}^{(T)}$

// geometry-aware scoring

Compute $\text{score}_{i,c}$ for each $x_i^u\in\mathcal{U}$ and $c\in\mathcal{Y}$ using $\mathcal{P}$, m, λ_κ, and α

// GeoDNI-based hidden-negative initialization

$\mathcal{A}_{\text{neg}}\leftarrow\{x_i^u\in\mathcal{U}:\text{score}_{i,c_{\text{neg}}}\text{ belongs to the top-}\tau\text{ fraction}\}$

$\gamma\leftarrow\min\{32,\lceil 0.02|\mathcal{U}|\rceil\}$

$\mathcal{N}^{(0)}\leftarrow\varnothing$

for $q\leftarrow 1$ **to** γ **do**

$$x_{(q)}^u\leftarrow\arg\max_{x_i^u\in\mathcal{A}_{\text{neg}}\setminus\mathcal{N}^{(q-1)}}\left[(1-\lambda_{\text{div}})R\big(\text{score}_{i,c_{\text{neg}}}\big)+\lambda_{\text{div}}\text{div}^{(q-1)}(x_i^u)\right]$$

$$\mathcal{N}^{(q)}\leftarrow\mathcal{N}^{(q-1)}\cup\{x_{(q)}^u\}$$

$\mathcal{C}^{(0)}\leftarrow\{(x_i^u,c_{\text{neg}}):x_i^u\in\mathcal{N}^{(\gamma)}\}$

$\mathcal{D}^{(0)}\leftarrow\mathcal{P}\cup\mathcal{C}^{(0)}$

for $t\leftarrow 0$ **to** $T-1$ **do**

// OOF TabPFN teacher prediction

Obtain F-fold OOF TabPFN predictions $\mathbf{p}_i^{(t)}$ for each $x_i^u\in\mathcal{U}$ using $\mathcal{D}^{(t)}$

// FGCR student prediction

Obtain FGCR predictions $\hat{\mathbf{p}}_i^{(t)}$ for each $x_i^u\in\mathcal{U}$

// confidence–geometry weighted distillation

$$w_i^{(t)}\leftarrow\left(p_i^{(t)}[1]-p_i^{(t)}[2]\right)\left[\lambda_{\text{score}}+(1-\lambda_{\text{score}})\textstyle\sum_{c\in\mathcal{Y}}p_{i,c}^{(t)}\text{score}_{i,c}\right]$$

$$\mathcal{L}_{\text{distilled}}^{(t)}\leftarrow\frac{\sum_{x_i^u\in\mathcal{U}}w_i^{(t)}\,\text{KL}\left(\mathbf{p}_i^{(t)}\,\middle\|\,\hat{\mathbf{p}}_i^{(t)}\right)}{\sum_{x_i^u\in\mathcal{U}}w_i^{(t)}+\epsilon}$$

// global geometry feedback

$$\mathcal{L}_{\text{ggeo}}^{(t)}\leftarrow-\frac{1}{|\mathcal{U}|}\textstyle\sum_{x_i^u\in\mathcal{U}}\sum_{c\in\mathcal{Y}}\text{score}_{i,c}\log\hat{p}_{i,c}^{(t)}$$

// selected-context geometry feedback

Construct $\mathbf{v}_i^{(t)}$ for samples in $\mathcal{C}^{(t)}$ according to the selected-context geometry target

$$\mathcal{L}_{\text{scgeo}}^{(t)}\leftarrow-\frac{1}{|\mathcal{C}^{(t)}|}\textstyle\sum_{(x_i^u,\cdot)\in\mathcal{C}^{(t)}}\sum_{c\in\mathcal{Y}}v_{i,c}^{(t)}\log\hat{p}_{i,c}^{(t)}$$

$$\mathcal{L}_{\text{geo}}^{(t)}\leftarrow\lambda_{\text{ggeo}}\mathcal{L}_{\text{ggeo}}^{(t)}+\lambda_{\text{scgeo}}\mathcal{L}_{\text{scgeo}}^{(t)}$$

$$\mathcal{L}_{FGCR}^{(t)}\leftarrow\lambda_{\text{distilled}}\mathcal{L}_{\text{distilled}}^{(t)}+\lambda_{\text{geo}}\mathcal{L}_{\text{geo}}^{(t)}$$

// FGCR parameter update

Update the FGCR parameters for E epochs by minimizing $\mathcal{L}_{FGCR}^{(t)}$

// pseudo-context retrieval and update

$$\text{rescore}_{i,c}^{(t)}\leftarrow\lambda_{\hat{p}}R\left(\hat{p}_{i,c}^{(t)}\right)+\lambda_{\text{diff}}R\left(\hat{p}_{i,c}^{(t)}-\max_{c'\neq c}\hat{p}_{i,c'}^{(t)}\right)+\lambda_pR\left(p_{i,c}^{(t)}\right)$$

// assign each unlabeled sample to its highest-rescore class

$$\tilde{y}_i^{(t+1)}\leftarrow\arg\max_{c\in\mathcal{Y}}\text{rescore}_{i,c}^{(t)}\text{ for each }x_i^u\in\mathcal{U}$$

$$\mathcal{C}^{(t+1)}\leftarrow\bigcup_{c\in\mathcal{Y}}\left\{(x_i^u,c):i\in\text{Top}_{\xi_c}\left(\{j:\tilde{y}_j^{(t+1)}=c\},\text{rescore}_{j,c}^{(t)}\right)\right\}$$

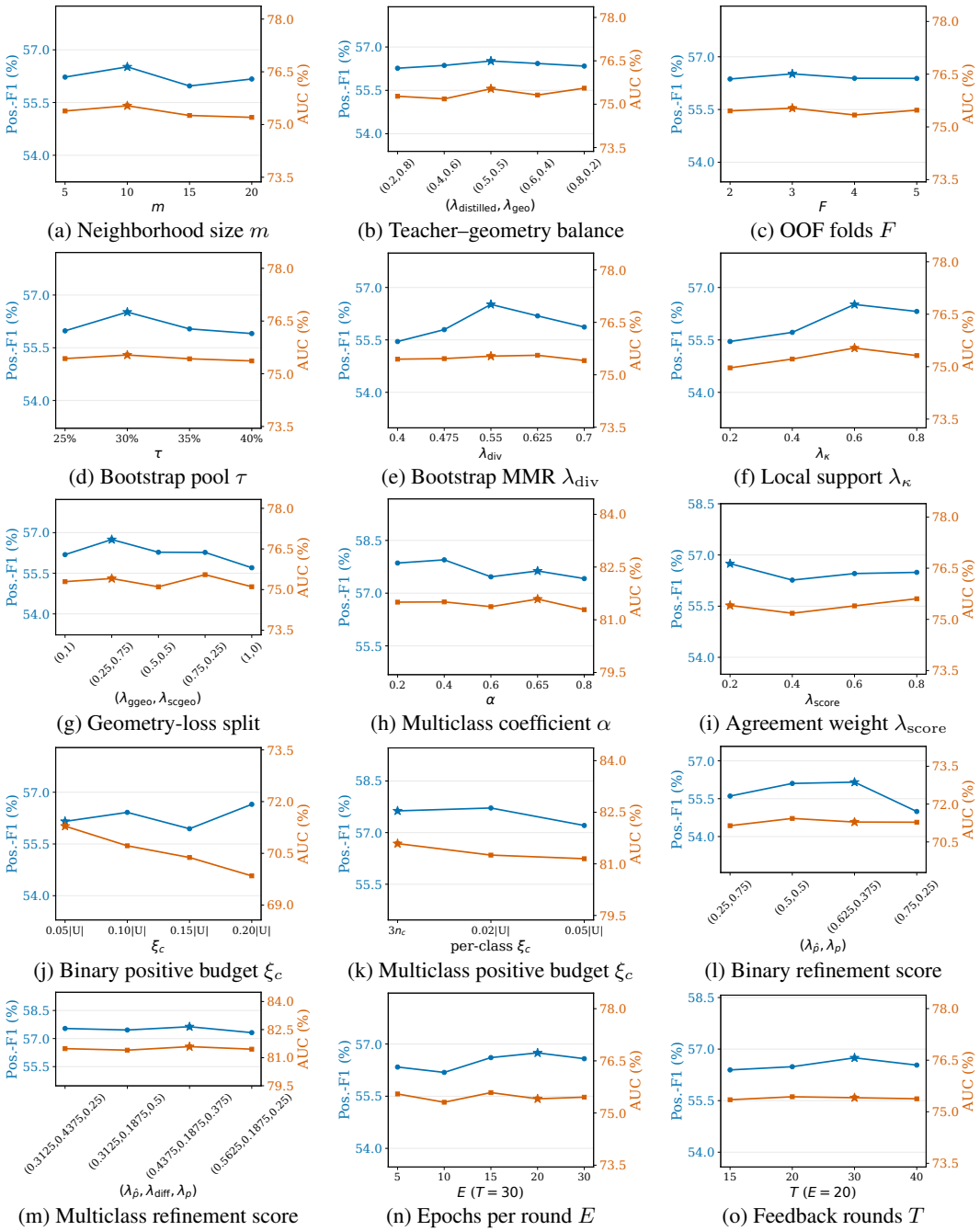
$$\mathcal{D}^{(t+1)}\leftarrow\mathcal{P}\cup\mathcal{C}^{(t+1)}$$

return $\mathcal{D}^{(T)}$

// use $\mathcal{D}^{(T)}$ as the TabPFN context for subsequent query prediction

E PARAMETER-SENSITIVITY ANALYSIS

Figure 3 presents the parameter-sensitivity results of DRIP-MPU. Overall, varying one parameter at a time leads to only modest fluctuations in both AUC and Pos.-F1, indicating that the framework is robust to the tested hyperparameter choices. Based on these results, we provide the recommended parameter settings in Appendix B.



(a) Neighborhood size m (b) Teacher–geometry balance (c) OOF folds F

(d) Bootstrap pool τ (e) Bootstrap MMR λ_{div} (f) Local support λ_κ

(g) Geometry-loss split (h) Multiclass coefficient α (i) Agreement weight λ_{score}

(j) Binary positive budget ξ_c (k) Multiclass positive budget ξ_c (l) Binary refinement score

(m) Multiclass refinement score (n) Epochs per round E (o) Feedback rounds T

Figure 3: Parameter-sensitivity analysis in the 5-shot setting. Blue lines denote Pos.-F1, and orange lines denote AUC. Stars mark the default setting for the corresponding metric. The left and right vertical axes report Pos.-F1 and AUC, respectively.

F COMPLETE RESULTS

Complete AUC and Pos.-F1 results across all 25 datasets. Bold and underlining indicate the best results in the fair comparison and in the comparison including oracle methods, respectively.

Table 6: Pos.-F1 (%) results on all 25 datasets in the 5-shot setting.

Method	Australian	blood-transfusion	climate-model sim.-crashes	credit-approval	credit-g	cylinder-bands	diabetes	dresses-sales	kc2
DRIP-MPU (Ours)	**74.9** ± 8.4	**46.6** ± 6.7	51.1 ± 5.8	**67.5** ± 20.6	**40.5** ± 11.1	**49.9** ± 9.7	**54.9** ± 5.5	44.4 ± 6.5	**46.4** ± 3.8
BBE-PU	0.3 ± 1.0	15.8 ± 21.1	0.4 ± 0.8	3.5 ± 3.7	0.0 ± 0.0	1.3 ± 2.8	2.8 ± 2.8	1.8 ± 3.2	14.6 ± 12.0
LaGAM-PU	35.3 ± 16.6	29.6 ± 18.9	32.8 ± 4.2	47.6 ± 11.0	24.3 ± 5.0	33.4 ± 11.6	35.7 ± 6.9	25.1 ± 5.5	37.1 ± 9.6
CGenPU	53.4 ± 12.5	39.7 ± 7.8	41.3 ± 18.9	49.2 ± 15.0	32.2 ± 11.4	38.2 ± 18.2	48.0 ± 8.5	24.1 ± 12.1	31.2 ± 11.6
VPU	0.6 ± 2.0	11.0 ± 12.8	0.4 ± 0.8	1.0 ± 1.6	0.3 ± 1.0	0.4 ± 1.3	1.1 ± 1.7	0.0 ± 0.0	0.9 ± 2.7
PAN	0.0 ± 0.0	35.4 ± 20.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	28.6 ± 20.4
HolisticPU	37.9 ± 9.5	21.4 ± 4.7	**74.3** ± 5.2	37.5 ± 10.2	37.7 ± 4.9	48.5 ± 8.1	37.7 ± 5.8	**50.4** ± 4.7	17.4 ± 6.7
P3Mix-C	3.5 ± 3.1	24.1 ± 8.2	0.4 ± 0.8	6.6 ± 3.4	1.6 ± 2.3	4.9 ± 5.3	5.8 ± 5.5	2.2 ± 3.1	9.5 ± 7.2
P3Mix-E	3.1 ± 3.3	19.9 ± 6.5	0.4 ± 0.8	5.0 ± 3.9	1.6 ± 1.7	4.9 ± 5.6	3.5 ± 2.8	0.9 ± 2.8	4.0 ± 5.7
PUET (KM2)	17.7 ± 23.8	8.5 ± 8.1	66.4 ± 44.9	37.9 ± 34.1	11.2 ± 19.0	33.5 ± 28.7	4.0 ± 7.9	21.1 ± 26.5	23.9 ± 14.9
ScalePU (KM2)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	1.0 ± 1.6	0.7 ± 1.4	0.4 ± 1.3	0.0 ± 0.0	0.5 ± 1.5	2.2 ± 4.9
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	65.5 ± 45.4	13.7 ± 27.6	0.0 ± 0.0	30.2 ± 29.0	0.0 ± 0.0	17.2 ± 27.8	3.3 ± 10.5
Dist-PU (KM2)	12.6 ± 25.7	0.0 ± 0.0	68.9 ± 43.2	32.7 ± 30.8	13.5 ± 18.5	36.3 ± 30.3	6.1 ± 13.0	21.7 ± 27.2	14.1 ± 19.9
PULDA (KM2)	14.8 ± 24.8	1.5 ± 4.6	67.4 ± 45.6	33.5 ± 29.5	10.2 ± 19.0	37.5 ± 28.9	5.1 ± 5.3	18.2 ± 28.3	8.3 ± 9.1
UREA (KM2)	32.3 ± 21.2	1.9 ± 6.0	43.5 ± 25.4	38.9 ± 22.5	12.9 ± 10.3	27.1 ± 16.8	18.2 ± 10.6	16.8 ± 16.8	26.2 ± 13.0
AREA (KM2)	1.6 ± 2.2	0.0 ± 0.0	0.4 ± 0.8	2.2 ± 2.2	1.0 ± 1.6	5.0 ± 4.2	0.4 ± 1.1	0.9 ± 1.9	4.2 ± 5.9
PUET (real)	66.1 ± 11.8	24.3 ± 11.0	95.4 ± 0.3	61.3 ± 20.4	21.5 ± 6.3	38.9 ± 18.6	34.0 ± 4.8	29.2 ± 5.7	37.2 ± 9.5
ScalePU (real)	0.3 ± 1.0	27.0 ± 21.9	0.2 ± 0.6	2.2 ± 2.2	0.3 ± 1.0	0.4 ± 1.3	1.8 ± 2.5	0.5 ± 1.5	9.1 ± 12.2
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	95.7 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	12.1 ± 13.1	6.5 ± 13.9	9.6 ± 8.9	18.9 ± 24.6
Dist-PU (real)	63.3 ± 12.4	18.6 ± 17.0	95.7 ± 0.0	56.8 ± 16.9	21.1 ± 5.0	45.8 ± 12.5	44.5 ± 8.2	33.7 ± 7.6	38.7 ± 8.4
PULDA (real)	12.7 ± 15.1	32.2 ± 16.3	95.7 ± 0.0	11.2 ± 7.7	2.8 ± 4.0	14.0 ± 24.5	17.5 ± 15.6	3.6 ± 4.6	15.0 ± 11.7
UREA (real)	58.0 ± 9.6	34.4 ± 15.1	61.7 ± 10.7	54.4 ± 15.5	28.1 ± 9.5	43.2 ± 11.1	42.0 ± 9.2	29.2 ± 10.6	40.4 ± 7.8
AREA (real)	2.2 ± 3.0	13.2 ± 10.0	1.0 ± 1.4	2.6 ± 2.5	2.3 ± 3.0	3.4 ± 3.3	2.5 ± 3.3	1.4 ± 3.1	1.7 ± 3.5

	pc3	segment	sonar	spambase	tic-tac-toe	wdbc	analcatdata_author	analcatdata_dmft	car
DRIP-MPU (Ours)	**29.2** ± 7.8	60.4 ± 25.4	**66.8** ± 8.8	**68.8** ± 8.4	49.9 ± 6.4	**91.2** ± 3.4	**95.3** ± 3.9	17.8 ± 3.1	**43.8** ± 6.6
BBE-PU	0.0 ± 0.0	0.0 ± 0.0	4.1 ± 7.9	0.0 ± 0.0	0.2 ± 0.5	9.8 ± 14.1	0.0 ± 0.0	6.1 ± 3.9	0.0 ± 0.0
LaGAM-PU	20.4 ± 7.6	50.4 ± 13.7	56.3 ± 14.3	25.6 ± 7.2	29.5 ± 7.8	61.3 ± 11.9	67.9 ± 3.0	**18.5** ± 2.0	20.8 ± 5.7
CGenPU	14.9 ± 9.0	40.3 ± 16.6	62.8 ± 17.0	46.1 ± 17.5	37.3 ± 14.2	59.7 ± 18.5	25.5 ± 7.2	17.0 ± 4.2	25.5 ± 7.9
VPU	0.0 ± 0.0	7.8 ± 11.4	0.0 ± 0.0	1.3 ± 1.1	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 1.8	6.0 ± 1.8	7.1 ± 6.5
PAN	0.0 ± 0.0	0.0 ± 0.0	36.6 ± 34.5	0.0 ± 0.0	0.0 ± 0.0	27.6 ± 40.9	0.0 ± 0.0	10.3 ± 2.9	0.0 ± 0.0
HolisticPU	10.6 ± 4.5	2.5 ± 1.3	37.5 ± 8.8	1.4 ± 1.1	**64.9** ± 5.5	16.5 ± 8.0	3.1 ± 1.6	14.7 ± 3.8	7.5 ± 1.8
P3Mix-C	4.7 ± 5.6	36.3 ± 9.5	5.9 ± 5.7	4.4 ± 2.4	2.5 ± 2.7	7.7 ± 4.7	3.3 ± 3.7	10.4 ± 2.9	30.8 ± 9.3
P3Mix-E	2.9 ± 4.1	33.8 ± 7.2	5.0 ± 7.8	3.5 ± 2.4	0.6 ± 1.1	4.5 ± 5.5	3.5 ± 4.4	9.1 ± 2.1	27.2 ± 9.9
PUET (KM2)	3.5 ± 5.6	48.6 ± 15.6	35.7 ± 32.7	13.1 ± 23.5	42.4 ± 32.6	8.3 ± 14.5	27.5 ± 12.4	4.1 ± 1.7	29.3 ± 5.7
ScalePU (KM2)	0.0 ± 0.0	7.1 ± 5.7	0.9 ± 2.7	1.0 ± 1.2	0.5 ± 0.8	0.0 ± 0.0	1.8 ± 5.6	0.1 ± 0.4	6.1 ± 5.6
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	27.6 ± 35.6	0.0 ± 0.0	28.6 ± 32.2	0.5 ± 1.5	3.0 ± 9.6	0.0 ± 0.0	0.0 ± 0.0
Dist-PU (KM2)	5.3 ± 8.7	43.6 ± 32.0	31.9 ± 33.4	12.2 ± 23.4	43.7 ± 36.3	10.6 ± 20.0	55.8 ± 19.9	0.0 ± 0.0	11.2 ± 10.3
PULDA (KM2)	1.2 ± 3.7	**76.9** ± 12.6	37.2 ± 28.4	13.2 ± 22.9	40.4 ± 40.6	2.3 ± 3.2	3.4 ± 5.5	1.7 ± 1.8	17.7 ± 12.3
UREA (KM2)	16.5 ± 9.8	51.2 ± 15.9	56.1 ± 12.7	26.3 ± 22.3	37.4 ± 17.8	37.2 ± 12.2	14.1 ± 5.9	2.5 ± 3.2	20.1 ± 3.5
AREA (KM2)	1.7 ± 5.3	7.2 ± 4.4	5.8 ± 8.6	1.3 ± 1.3	0.6 ± 0.8	3.3 ± 2.2	2.7 ± 2.9	1.9 ± 2.6	13.1 ± 5.4
PUET (real)	3.5 ± 5.3	77.2 ± 7.0	66.5 ± 10.5	34.9 ± 15.9	62.1 ± 6.6	79.1 ± 5.6	51.9 ± 6.8	10.8 ± 2.3	37.7 ± 6.9
ScalePU (real)	0.6 ± 1.9	9.0 ± 6.9	1.7 ± 3.7	1.1 ± 1.1	0.5 ± 0.8	0.9 ± 2.0	1.8 ± 5.6	7.7 ± 3.9	6.9 ± 2.4
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	54.6 ± 10.2	0.0 ± 0.0	28.0 ± 15.0	70.4 ± 14.0	2.7 ± 4.7	0.0 ± 0.0	0.0 ± 0.0
Dist-PU (real)	8.8 ± 8.8	82.1 ± 7.3	70.5 ± 7.9	15.2 ± 9.1	75.1 ± 4.8	77.7 ± 6.4	78.7 ± 11.1	4.7 ± 2.4	13.5 ± 4.9
PULDA (real)	4.9 ± 7.0	84.8 ± 5.8	68.8 ± 0.0	8.0 ± 7.3	78.9 ± 0.0	1.4 ± 2.2	2.2 ± 2.8	14.6 ± 4.4	17.8 ± 10.0
UREA (real)	18.9 ± 10.1	74.7 ± 7.4	66.7 ± 13.6	46.8 ± 6.5	45.3 ± 8.1	66.1 ± 11.6	19.6 ± 8.1	13.1 ± 2.8	16.8 ± 4.2
AREA (real)	2.2 ± 6.8	14.2 ± 11.4	2.5 ± 5.7	1.3 ± 1.1	0.8 ± 0.8	1.8 ± 3.2	4.0 ± 3.7	8.8 ± 3.3	16.5 ± 7.6

	Gesture Phase Seg. Proc.	first-order theorem prov.	mfeat-karhunen	mfeat-fourier	MiceProtein	satimage	steel-plates fault	Average Pos.-F1 (%)	Average Rank	Wins
DRIP-MPU (Ours)	25.9 ± 3.2	**18.6** ± 2.6	**82.4** ± 3.8	**69.0** ± 3.7	**93.8** ± 5.8	**72.1** ± 2.4	**57.6** ± 5.9	**56.7** ± 0.9	**1.400**	**19**
BBE-PU	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	2.4 ± 1.3	13.040	0
LaGAM-PU	**28.4** ± 1.9	17.1 ± 4.2	68.4 ± 3.7	47.3 ± 5.4	59.5 ± 6.3	54.3 ± 7.1	44.6 ± 3.5	38.8 ± 1.7	3.400	2
CGenPU	18.3 ± 2.7	14.4 ± 3.0	11.4 ± 4.3	8.7 ± 1.5	12.9 ± 5.6	19.2 ± 6.1	39.2 ± 6.3	32.4 ± 2.5	3.840	0
VPU	0.0 ± 0.1	0.6 ± 0.3	0.2 ± 0.4	0.2 ± 0.3	0.3 ± 0.6	0.2 ± 0.4	1.9 ± 1.5	1.7 ± 0.9	13.020	0
PAN	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	5.5 ± 2.1	13.100	0
HolisticPU	10.0 ± 2.7	12.4 ± 1.7	1.9 ± 0.6	2.9 ± 0.9	1.7 ± 1.1	1.6 ± 1.3	3.9 ± 1.4	22.2 ± 0.8	6.320	3
P3Mix-C	3.9 ± 1.1	10.6 ± 1.8	3.0 ± 1.4	3.7 ± 1.1	8.7 ± 3.1	10.9 ± 3.7	25.5 ± 6.5	9.2 ± 1.1	8.160	0
P3Mix-E	2.7 ± 0.7	9.9 ± 1.5	2.7 ± 1.2	2.8 ± 1.2	6.2 ± 2.6	10.8 ± 2.2	25.7 ± 6.0	7.8 ± 0.7	9.400	0
PUET (KM2)	6.7 ± 5.0	7.0 ± 3.1	7.8 ± 3.0	8.8 ± 3.2	39.7 ± 11.3	45.0 ± 5.4	28.1 ± 5.5	23.2 ± 2.6	6.080	0
ScalePU (KM2)	0.2 ± 0.1	2.9 ± 1.9	0.0 ± 0.0	0.1 ± 0.2	0.3 ± 0.7	4.3 ± 5.4	9.4 ± 6.0	1.6 ± 0.7	13.400	0
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	7.6 ± 2.7	13.060	0
Dist-PU (KM2)	5.9 ± 5.1	2.1 ± 3.2	27.5 ± 6.3	16.9 ± 4.5	41.7 ± 27.1	18.2 ± 6.8	19.8 ± 6.4	22.1 ± 2.7	6.840	0
PULDA (KM2)	5.9 ± 5.3	7.5 ± 3.2	2.6 ± 1.0	3.3 ± 1.0	33.6 ± 21.1	17.6 ± 5.9	32.9 ± 8.8	19.7 ± 2.5	7.440	1
UREA (KM2)	7.9 ± 1.4	7.9 ± 4.3	5.1 ± 2.0	5.1 ± 2.3	4.6 ± 1.5	35.4 ± 8.9	13.3 ± 7.3	22.3 ± 1.7	6.160	0
AREA (KM2)	0.3 ± 0.5	2.9 ± 1.4	2.5 ± 1.0	1.3 ± 0.6	0.3 ± 0.5	6.6 ± 2.5	10.2 ± 5.8	3.1 ± 0.9	11.340	0
PUET (real)	4.0 ± 1.6	11.3 ± 2.1	9.8 ± 2.2	9.4 ± 2.3	30.5 ± 3.1	65.4 ± 6.5	46.0 ± 8.1	40.3 ± 0.9	–	–
ScalePU (real)	0.2 ± 0.1	3.3 ± 0.6	0.0 ± 0.0	0.1 ± 0.2	0.0 ± 0.0	3.9 ± 3.7	13.5 ± 6.8	3.7 ± 0.9	–	–
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	11.9 ± 1.7	–	–
Dist-PU (real)	1.7 ± 0.9	8.2 ± 2.9	34.2 ± 5.2	20.4 ± 4.4	58.6 ± 5.4	43.6 ± 4.3	38.4 ± 8.2	42.0 ± 1.5	–	–
PULDA (real)	0.8 ± 0.4	13.4 ± 3.5	3.4 ± 1.4	3.3 ± 0.9	32.8 ± 5.1	36.3 ± 7.9	38.1 ± 8.7	24.6 ± 1.2	–	–
UREA (real)	9.3 ± 1.4	10.1 ± 1.4	6.3 ± 3.4	5.1 ± 2.2	7.0 ± 2.6	45.9 ± 11.2	16.2 ± 5.3	34.4 ± 2.0	–	–
AREA (real)	0.6 ± 0.4	4.9 ± 1.3	2.7 ± 0.9	2.0 ± 1.0	0.0 ± 0.0	8.7 ± 3.3	14.5 ± 6.7	4.6 ± 1.0	–	–

Table 7: AUC (%) results on all 25 datasets in the 5-shot setting.

Method	Australian	blood-transfusion	climate-model sim.-crashes	credit-approval	credit-g	cylinder-bands	diabetes	dresses-sales	kc2
DRIP-MPU (Ours)	**85.5** ± 7.5	**68.0** ± 6.4	45.2 ± 12.1	**79.0** ± 18.7	**59.9** ± 8.6	**59.5** ± 10.0	**69.5** ± 8.8	52.3 ± 5.7	71.6 ± 5.1
BBE-PU	70.8 ± 10.2	66.8 ± 6.4	53.0 ± 13.3	72.9 ± 12.4	53.1 ± 6.4	53.5 ± 10.8	65.0 ± 9.1	**53.8** ± 4.5	70.3 ± 3.4
LaGAM-PU	63.4 ± 16.9	53.6 ± 14.2	50.8 ± 14.3	70.0 ± 14.5	54.3 ± 5.1	55.9 ± 11.1	64.9 ± 7.5	53.5 ± 5.4	59.9 ± 15.8
CGenPU	72.7 ± 8.9	59.4 ± 12.1	45.8 ± 13.9	71.5 ± 13.4	55.4 ± 10.7	54.0 ± 11.1	64.4 ± 7.5	52.1 ± 3.3	56.9 ± 12.4
VPU	54.9 ± 5.2	59.0 ± 5.9	**57.0** ± 10.5	64.4 ± 7.5	49.9 ± 5.6	54.4 ± 2.5	57.7 ± 5.3	52.7 ± 5.2	54.0 ± 4.7
PAN	49.5 ± 5.4	67.4 ± 8.4	53.7 ± 6.2	62.9 ± 7.5	43.9 ± 4.5	37.4 ± 4.7	55.4 ± 4.1	53.6 ± 5.0	**75.3** ± 0.8
HolisticPU	29.2 ± 6.4	39.8 ± 8.6	45.5 ± 10.5	28.7 ± 12.7	46.7 ± 6.1	50.4 ± 9.7	36.0 ± 10.5	47.8 ± 3.7	35.5 ± 8.0
P3Mix-C	59.8 ± 7.7	60.7 ± 3.8	52.1 ± 12.4	68.1 ± 11.2	54.6 ± 4.6	53.1 ± 5.0	64.4 ± 7.4	53.0 ± 4.4	65.8 ± 5.9
P3Mix-E	59.9 ± 7.8	62.7 ± 5.2	54.0 ± 10.9	67.1 ± 10.7	53.6 ± 4.1	48.0 ± 4.7	63.5 ± 7.1	52.9 ± 5.1	55.4 ± 7.9
PUET (KM2)	69.7 ± 10.5	56.3 ± 4.8	51.0 ± 7.1	71.2 ± 14.2	59.4 ± 4.5	56.3 ± 10.6	62.0 ± 6.2	52.0 ± 4.1	68.1 ± 4.8
ScalePU (KM2)	64.3 ± 7.1	64.0 ± 9.7	48.9 ± 9.2	69.5 ± 11.3	53.3 ± 5.7	57.8 ± 6.9	64.4 ± 9.1	52.8 ± 4.4	67.2 ± 6.7
nnPU (KM2)	53.8 ± 12.1	52.9 ± 2.9	51.8 ± 12.8	66.6 ± 13.7	46.8 ± 7.4	50.7 ± 17.5	46.7 ± 3.9	52.8 ± 5.8	63.4 ± 10.8
Dist-PU (KM2)	74.3 ± 8.4	60.2 ± 5.0	44.6 ± 9.5	68.2 ± 17.9	50.1 ± 8.8	55.0 ± 17.0	54.2 ± 12.8	51.2 ± 4.8	70.9 ± 5.1
PULDA (KM2)	70.0 ± 8.9	65.7 ± 3.7	46.5 ± 9.5	64.8 ± 12.8	55.4 ± 5.3	59.5 ± 6.7	63.0 ± 8.0	50.8 ± 6.5	62.2 ± 8.6
UREA (KM2)	57.7 ± 5.6	58.6 ± 8.6	49.3 ± 11.1	58.3 ± 9.9	54.8 ± 4.2	57.9 ± 5.7	54.0 ± 2.8	51.8 ± 2.9	51.9 ± 9.5
AREA (KM2)	55.5 ± 12.0	57.3 ± 8.2	50.1 ± 14.8	61.4 ± 15.0	52.8 ± 5.4	59.3 ± 5.1	56.3 ± 8.8	51.0 ± 3.6	55.5 ± 10.4
PUET (real)	78.9 ± 9.0	62.3 ± 6.1	48.5 ± 14.0	77.0 ± 15.0	59.4 ± 7.3	63.8 ± 13.1	70.1 ± 5.2	52.5 ± 6.1	71.5 ± 1.9
ScalePU (real)	64.0 ± 7.0	64.3 ± 9.8	49.2 ± 8.9	69.1 ± 11.3	53.4 ± 5.9	57.7 ± 7.4	64.3 ± 9.2	52.8 ± 4.3	67.0 ± 6.3
nnPU (real)	71.5 ± 6.9	61.6 ± 4.6	50.2 ± 12.4	71.5 ± 8.3	48.3 ± 4.3	53.2 ± 15.1	63.3 ± 8.5	53.6 ± 4.6	73.0 ± 1.8
Dist-PU (real)	77.0 ± 9.7	68.5 ± 5.2	43.8 ± 5.9	72.2 ± 13.2	52.5 ± 6.9	55.9 ± 12.4	65.9 ± 9.3	53.7 ± 4.9	73.6 ± 2.0
PULDA (real)	70.3 ± 12.6	65.1 ± 4.8	45.6 ± 7.3	69.1 ± 16.4	54.2 ± 4.0	55.1 ± 10.8	61.6 ± 8.4	52.5 ± 5.5	64.0 ± 8.4
UREA (real)	65.5 ± 6.7	63.0 ± 7.6	52.0 ± 12.3	65.0 ± 11.1	54.2 ± 3.0	57.6 ± 6.1	59.8 ± 4.1	51.3 ± 4.1	61.2 ± 4.3
AREA (real)	57.9 ± 11.1	58.1 ± 9.0	51.4 ± 12.8	64.2 ± 13.3	52.6 ± 5.9	59.8 ± 5.2	57.8 ± 10.6	51.4 ± 4.0	58.0 ± 10.8

	pc3	segment	sonar	spambase	tic-tac-toe	wdbc	analcatdata_author	analcatdata_dmft	car
DRIP-MPU (Ours)	71.5 ± 12.1	97.4 ± 3.8	**72.4** ± 9.6	**84.8** ± 6.8	53.8 ± 6.5	**99.0** ± 1.1	**99.1** ± 1.6	51.1 ± 2.3	**86.9** ± 2.7
BBE-PU	68.9 ± 8.5	92.4 ± 4.9	70.4 ± 7.4	53.6 ± 6.3	52.6 ± 6.0	93.0 ± 3.3	92.6 ± 2.4	50.5 ± 3.2	78.0 ± 4.9
LaGAM-PU	58.0 ± 10.2	85.4 ± 8.6	68.6 ± 15.1	44.1 ± 3.8	51.3 ± 7.4	88.2 ± 7.7	95.2 ± 1.6	51.4 ± 2.6	79.8 ± 4.3
CGenPU	49.2 ± 13.1	76.3 ± 17.4	70.1 ± 11.5	54.8 ± 20.0	51.0 ± 6.5	80.6 ± 20.8	58.5 ± 9.7	**52.0** ± 3.4	71.4 ± 7.2
VPU	58.6 ± 6.7	88.9 ± 3.5	64.9 ± 10.1	62.0 ± 3.5	52.7 ± 7.1	70.7 ± 1.6	67.2 ± 2.2	50.0 ± 2.2	79.1 ± 4.0
PAN	**73.1** ± 2.9	82.9 ± 9.6	52.9 ± 12.7	54.3 ± 6.6	50.7 ± 8.5	80.8 ± 19.4	60.1 ± 2.8	51.1 ± 3.2	66.7 ± 2.1
HolisticPU	34.3 ± 11.6	23.4 ± 3.9	31.3 ± 7.5	56.0 ± 3.8	48.4 ± 6.2	9.4 ± 4.8	10.2 ± 3.5	48.5 ± 1.8	22.9 ± 3.5
P3Mix-C	64.6 ± 11.3	95.7 ± 2.0	65.7 ± 11.5	63.3 ± 3.8	**55.0** ± 6.2	84.0 ± 4.6	69.9 ± 2.9	50.4 ± 2.6	78.9 ± 4.3
P3Mix-E	55.0 ± 7.9	95.6 ± 2.7	64.5 ± 10.2	53.0 ± 4.8	54.2 ± 5.8	76.6 ± 7.1	61.8 ± 3.2	50.4 ± 2.8	74.9 ± 4.8
PUET (KM2)	65.9 ± 6.8	94.6 ± 5.8	68.6 ± 8.6	66.7 ± 8.7	53.5 ± 5.8	87.9 ± 4.8	95.1 ± 2.3	50.1 ± 1.3	71.6 ± 5.7
ScalePU (KM2)	61.1 ± 7.4	92.2 ± 4.2	69.8 ± 9.2	52.2 ± 7.5	54.1 ± 6.5	86.9 ± 7.5	79.3 ± 2.6	51.4 ± 3.2	72.9 ± 5.1
nnPU (KM2)	60.6 ± 3.8	73.2 ± 2.5	69.6 ± 6.8	51.9 ± 1.6	51.9 ± 7.3	71.6 ± 13.7	85.9 ± 5.6	48.2 ± 1.7	58.8 ± 5.6
Dist-PU (KM2)	65.4 ± 10.0	93.9 ± 3.7	66.3 ± 11.6	55.9 ± 9.5	52.0 ± 5.9	88.3 ± 11.3	96.5 ± 3.8	48.5 ± 1.1	65.5 ± 8.1
PULDA (KM2)	61.1 ± 13.5	**98.1** ± 1.0	64.3 ± 13.0	54.2 ± 8.2	52.9 ± 4.5	78.9 ± 8.0	86.5 ± 2.9	50.0 ± 2.4	81.7 ± 4.5
UREA (KM2)	54.7 ± 4.5	84.3 ± 8.4	64.3 ± 7.5	53.7 ± 6.4	50.6 ± 3.5	62.0 ± 4.6	55.8 ± 2.5	51.2 ± 3.2	72.6 ± 8.7
AREA (KM2)	53.9 ± 8.4	90.1 ± 7.1	63.4 ± 8.2	45.8 ± 6.8	50.5 ± 3.6	71.5 ± 12.6	57.1 ± 7.5	51.7 ± 3.4	85.5 ± 5.1
PUET (real)	67.8 ± 9.1	96.8 ± 4.6	73.4 ± 8.7	81.4 ± 7.3	52.7 ± 5.8	98.1 ± 0.9	97.7 ± 1.0	50.5 ± 2.8	75.6 ± 4.0
ScalePU (real)	61.0 ± 7.0	92.0 ± 4.4	69.3 ± 9.3	50.5 ± 6.2	54.1 ± 6.4	86.4 ± 7.5	79.1 ± 3.0	51.5 ± 3.2	72.5 ± 5.5
nnPU (real)	63.4 ± 3.6	78.3 ± 2.7	71.6 ± 5.9	50.6 ± 2.3	53.5 ± 6.8	98.3 ± 1.4	80.1 ± 2.8	49.0 ± 1.9	56.7 ± 3.4
Dist-PU (real)	68.2 ± 12.4	94.9 ± 2.9	74.6 ± 8.2	42.4 ± 5.1	53.2 ± 5.5	96.0 ± 2.7	98.2 ± 0.5	49.9 ± 2.7	69.5 ± 8.6
PULDA (real)	61.1 ± 13.2	98.4 ± 0.9	49.2 ± 5.1	42.3 ± 7.5	53.3 ± 5.0	77.0 ± 10.5	89.4 ± 2.3	50.1 ± 3.1	85.3 ± 3.7
UREA (real)	55.7 ± 5.6	84.1 ± 5.3	70.2 ± 10.4	58.8 ± 3.0	50.9 ± 2.2	75.0 ± 7.2	59.1 ± 4.7	50.6 ± 2.3	61.1 ± 2.4
AREA (real)	52.6 ± 8.3	90.2 ± 5.9	63.0 ± 8.6	46.8 ± 7.7	52.2 ± 4.9	78.0 ± 11.6	60.5 ± 8.8	50.7 ± 3.1	86.3 ± 3.1

	Gesture Phase Seg. Proc.	first-order theorem prov.	mfeat-karhunen	mfeat-fourier	MiceProtein	satimage	steel-plates fault	Average AUC (%)	Average Rank	Wins
DRIP-MPU (Ours)	56.3 ± 3.6	**57.3** ± 2.7	**95.3** ± 3.1	**93.5** ± 2.3	**98.2** ± 2.1	**95.3** ± 0.6	82.9 ± 4.6	**75.4** ± 1.1	**2.600**	**16**
BBE-PU	55.8 ± 2.3	51.1 ± 1.7	90.5 ± 0.7	83.3 ± 2.1	89.8 ± 1.9	51.4 ± 7.6	83.6 ± 2.9	68.7 ± 1.2	5.800	1
LaGAM-PU	57.7 ± 2.4	56.5 ± 2.9	91.9 ± 1.7	85.6 ± 2.2	86.2 ± 2.6	88.7 ± 1.0	82.9 ± 1.7	67.9 ± 1.9	6.360	0
CGenPU	51.3 ± 3.3	53.0 ± 2.9	61.9 ± 5.2	56.5 ± 3.0	59.9 ± 3.9	66.8 ± 5.5	77.1 ± 2.4	60.9 ± 2.0	9.560	1
VPU	55.4 ± 1.5	55.7 ± 2.6	76.8 ± 1.8	70.6 ± 2.0	80.0 ± 1.8	77.8 ± 1.8	81.5 ± 2.6	63.8 ± 1.1	8.960	1
PAN	**59.3** ± 3.7	52.7 ± 1.8	62.2 ± 2.7	54.9 ± 1.5	66.3 ± 1.7	55.9 ± 5.2	69.5 ± 3.6	59.7 ± 1.8	9.840	3
HolisticPU	50.8 ± 1.4	48.8 ± 2.8	17.7 ± 0.8	21.3 ± 1.7	20.5 ± 3.3	17.6 ± 7.6	27.0 ± 3.9	33.9 ± 1.3	15.200	0
P3Mix-C	56.9 ± 2.4	54.5 ± 2.3	75.0 ± 2.4	69.7 ± 2.3	79.7 ± 2.1	82.6 ± 1.9	81.9 ± 2.8	66.4 ± 1.5	6.800	1
P3Mix-E	54.9 ± 3.5	54.2 ± 2.2	67.8 ± 1.9	62.7 ± 2.3	75.1 ± 3.2	79.5 ± 2.8	79.8 ± 3.7	63.1 ± 1.4	8.760	0
PUET (KM2)	54.3 ± 7.3	54.8 ± 2.3	89.8 ± 0.8	85.0 ± 1.5	92.7 ± 2.7	92.5 ± 1.1	82.3 ± 1.9	70.0 ± 1.2	6.000	0
ScalePU (KM2)	50.4 ± 2.3	53.8 ± 2.4	75.9 ± 2.3	71.9 ± 2.3	79.9 ± 2.4	80.2 ± 2.3	80.1 ± 2.2	66.2 ± 1.2	7.480	0
nnPU (KM2)	55.6 ± 3.0	51.5 ± 1.6	82.4 ± 1.7	76.4 ± 1.6	68.5 ± 7.3	53.0 ± 3.1	67.6 ± 4.6	60.5 ± 1.9	11.080	0
Dist-PU (KM2)	54.3 ± 5.7	53.8 ± 2.3	90.8 ± 1.0	84.2 ± 1.5	90.1 ± 3.7	92.4 ± 1.4	83.5 ± 2.8	68.4 ± 1.5	7.440	0
PULDA (KM2)	53.6 ± 5.5	54.9 ± 2.6	89.7 ± 1.4	80.7 ± 1.8	92.9 ± 2.7	85.4 ± 2.8	**85.2** ± 2.0	68.3 ± 1.4	6.880	2
UREA (KM2)	51.3 ± 2.2	53.4 ± 2.4	52.7 ± 1.4	53.6 ± 1.5	52.8 ± 1.2	68.1 ± 4.2	56.2 ± 3.5	57.3 ± 0.9	11.920	0
AREA (KM2)	41.8 ± 0.9	53.8 ± 2.8	59.5 ± 2.4	61.1 ± 3.5	59.9 ± 1.9	78.7 ± 4.1	67.7 ± 2.3	59.7 ± 1.5	11.320	0
PUET (real)	63.2 ± 2.7	56.1 ± 2.4	91.0 ± 1.4	84.9 ± 1.5	92.2 ± 2.1	94.3 ± 1.0	85.2 ± 2.1	73.8 ± 1.3	–	–
ScalePU (real)	50.4 ± 1.6	53.3 ± 1.5	75.9 ± 2.0	71.8 ± 2.3	80.0 ± 2.2	79.2 ± 1.6	79.8 ± 2.8	65.9 ± 1.3	–	–
nnPU (real)	55.7 ± 1.8	51.5 ± 1.6	84.1 ± 1.4	78.3 ± 1.9	83.2 ± 1.5	57.1 ± 2.8	74.9 ± 3.7	65.3 ± 1.3	–	–
Dist-PU (real)	48.4 ± 1.5	55.9 ± 2.8	90.9 ± 1.0	85.6 ± 1.2	93.3 ± 1.6	93.6 ± 0.8	86.9 ± 1.8	70.6 ± 1.0	–	–
PULDA (real)	46.1 ± 2.4	56.0 ± 3.3	90.6 ± 1.5	80.9 ± 2.0	95.8 ± 1.3	87.4 ± 1.5	86.2 ± 1.3	67.5 ± 1.4	–	–
UREA (real)	51.0 ± 1.9	52.4 ± 1.3	53.0 ± 1.4	55.1 ± 1.9	53.1 ± 1.8	73.2 ± 5.8	56.0 ± 2.6	59.6 ± 1.2	–	–
AREA (real)	42.3 ± 1.2	54.1 ± 3.6	59.1 ± 2.2	60.9 ± 2.7	60.4 ± 3.2	79.4 ± 2.8	67.0 ± 3.2	60.6 ± 1.1	–	–

Table 8: Pos.-F1 (%) results on all 25 datasets in the 10-shot setting.

Method	Australian	blood-transfusion	climate-model sim.-crashes	credit-approval	credit-g	cylinder-bands	diabetes	dresses-sales	kc2
DRIP-MPU (Ours)	**74.3** ± 10.0	44.2 ± 10.5	63.0 ± 7.8	**73.2** ± 5.7	**48.3** ± 5.8	**52.4** ± 12.6	**59.4** ± 5.2	45.4 ± 5.5	**43.5** ± 7.4
BBE-PU	13.7 ± 13.6	**46.3** ± 10.6	4.2 ± 4.7	20.8 ± 7.1	6.2 ± 5.6	5.7 ± 5.4	27.8 ± 7.9	4.3 ± 6.3	32.7 ± 10.9
LaGAM-PU	43.5 ± 17.9	40.5 ± 13.6	41.7 ± 7.8	48.4 ± 14.2	31.1 ± 5.2	33.1 ± 12.6	36.2 ± 8.1	25.4 ± 7.4	40.8 ± 8.4
CGenPU	59.4 ± 13.5	40.2 ± 6.9	60.4 ± 13.1	57.8 ± 8.2	38.1 ± 8.0	48.2 ± 11.5	51.8 ± 6.0	38.3 ± 8.9	35.7 ± 9.5
VPU	1.9 ± 2.2	25.8 ± 17.1	0.2 ± 0.6	2.9 ± 3.1	1.6 ± 2.3	2.9 ± 4.3	3.2 ± 3.5	0.9 ± 2.0	7.1 ± 9.5
PAN	34.5 ± 33.7	41.2 ± 6.6	47.8 ± 44.9	15.7 ± 29.2	0.0 ± 0.0	28.8 ± 27.1	30.9 ± 24.9	0.0 ± 0.0	38.1 ± 5.8
HolisticPU	35.0 ± 11.5	17.7 ± 7.0	66.6 ± 6.2	34.1 ± 6.7	32.1 ± 3.9	45.2 ± 6.8	34.8 ± 6.2	**49.8** ± 4.5	18.3 ± 6.7
P3Mix-C	2.8 ± 3.9	29.3 ± 11.2	0.8 ± 1.4	8.0 ± 3.7	4.8 ± 3.4	6.6 ± 4.7	7.5 ± 3.4	4.0 ± 3.7	13.7 ± 11.1
P3Mix-E	3.1 ± 3.8	25.3 ± 10.2	1.0 ± 1.9	6.8 ± 4.2	4.5 ± 2.7	6.1 ± 5.9	6.2 ± 4.1	4.4 ± 2.9	11.2 ± 12.3
PUET (KM2)	15.4 ± 18.0	14.0 ± 10.8	68.1 ± 44.4	32.3 ± 26.7	22.0 ± 18.4	17.9 ± 26.0	7.6 ± 10.0	23.3 ± 30.1	36.1 ± 11.4
ScalePU (KM2)	4.2 ± 5.9	0.0 ± 0.0	0.4 ± 0.8	5.3 ± 4.1	1.0 ± 2.2	0.0 ± 0.0	5.1 ± 9.7	0.9 ± 1.9	15.6 ± 22.0
nnPU (KM2)	3.3 ± 10.5	0.0 ± 0.0	66.9 ± 46.2	17.0 ± 28.4	13.0 ± 21.0	11.9 ± 25.2	0.0 ± 0.0	23.7 ± 30.6	23.0 ± 21.2
Dist-PU (KM2)	20.7 ± 22.9	3.9 ± 12.4	**70.7** ± 40.4	32.7 ± 24.2	26.5 ± 19.8	17.1 ± 26.8	9.9 ± 16.1	23.7 ± 30.5	33.5 ± 20.2
PULDA (KM2)	19.8 ± 16.0	7.4 ± 17.1	67.8 ± 44.9	26.2 ± 20.3	17.9 ± 20.3	16.9 ± 23.1	14.9 ± 13.3	23.7 ± 30.5	25.6 ± 16.8
UREA (KM2)	37.4 ± 16.8	11.3 ± 16.3	50.7 ± 30.7	41.6 ± 16.9	31.8 ± 16.6	34.0 ± 12.3	23.6 ± 6.4	15.8 ± 13.6	37.4 ± 11.1
AREA (KM2)	3.4 ± 3.6	7.9 ± 15.3	2.4 ± 1.8	6.6 ± 2.6	1.6 ± 2.3	8.1 ± 6.0	3.8 ± 3.8	2.7 ± 4.2	6.6 ± 3.5
PUET (real)	69.4 ± 14.7	29.3 ± 11.1	95.6 ± 0.2	69.4 ± 10.9	28.7 ± 9.0	48.1 ± 9.5	42.7 ± 9.0	31.8 ± 8.3	40.9 ± 5.8
ScalePU (real)	7.9 ± 6.1	15.3 ± 24.9	0.4 ± 0.8	6.6 ± 3.1	1.3 ± 2.7	0.9 ± 1.8	18.1 ± 10.2	2.3 ± 3.2	14.0 ± 23.0
nnPU (real)	36.2 ± 26.3	0.0 ± 0.0	95.7 ± 0.0	44.4 ± 27.9	0.0 ± 0.0	36.3 ± 19.1	36.2 ± 18.1	26.1 ± 8.1	26.4 ± 18.2
Dist-PU (real)	65.3 ± 15.3	19.1 ± 25.1	95.7 ± 0.0	66.7 ± 10.9	31.8 ± 6.8	47.6 ± 11.4	49.8 ± 6.4	35.6 ± 5.1	43.6 ± 5.6
PULDA (real)	55.6 ± 11.5	42.4 ± 16.9	95.7 ± 0.0	51.6 ± 15.5	5.6 ± 5.7	59.7 ± 0.0	33.9 ± 9.6	9.5 ± 5.6	31.4 ± 17.1
UREA (real)	61.5 ± 11.3	30.7 ± 14.5	74.2 ± 8.8	63.8 ± 10.8	33.1 ± 9.6	47.1 ± 7.5	49.4 ± 7.8	28.4 ± 7.7	38.8 ± 9.6
AREA (real)	4.1 ± 3.0	19.9 ± 14.5	2.6 ± 2.3	5.1 ± 3.7	1.9 ± 2.7	7.4 ± 3.7	6.0 ± 3.6	1.4 ± 2.2	7.1 ± 6.9

	pc3	segment	sonar	spambase	tic-tac-toe	wdbc	analcatdata_author	analcatdata_dmft	car
DRIP-MPU (Ours)	**28.9** ± 7.1	48.3 ± 15.5	**73.8** ± 8.5	**75.1** ± 5.8	57.0 ± 4.8	**91.3** ± 4.3	**97.9** ± 1.6	17.1 ± 2.1	**47.6** ± 6.4
BBE-PU	4.5 ± 7.7	0.0 ± 0.0	9.1 ± 9.6	0.0 ± 0.0	11.3 ± 9.4	28.0 ± 13.3	27.8 ± 15.2	14.7 ± 2.6	2.9 ± 5.0
LaGAM-PU	27.1 ± 9.6	63.6 ± 11.2	59.6 ± 12.5	29.5 ± 7.6	39.3 ± 7.5	55.1 ± 14.7	72.7 ± 4.2	**19.6** ± 3.2	24.1 ± 3.5
CGenPU	22.8 ± 8.6	49.3 ± 9.3	68.4 ± 8.2	49.9 ± 8.7	56.5 ± 5.9	72.2 ± 13.5	41.8 ± 12.6	15.9 ± 2.5	29.2 ± 7.7
VPU	0.6 ± 1.9	47.1 ± 21.2	4.2 ± 7.1	1.4 ± 1.3	0.9 ± 1.3	1.4 ± 2.2	2.9 ± 3.8	10.6 ± 2.5	29.4 ± 8.8
PAN	0.0 ± 0.0	0.0 ± 0.0	49.8 ± 29.1	0.0 ± 0.0	0.8 ± 2.4	39.9 ± 43.3	0.0 ± 0.0	16.3 ± 5.1	0.6 ± 1.8
HolisticPU	9.1 ± 3.4	3.6 ± 1.2	33.9 ± 7.9	2.7 ± 1.4	**61.1** ± 6.1	11.8 ± 7.8	2.1 ± 1.3	11.2 ± 3.7	6.8 ± 2.1
P3Mix-C	3.2 ± 3.6	40.7 ± 9.1	14.6 ± 7.2	5.7 ± 2.1	3.4 ± 2.4	11.8 ± 6.6	6.4 ± 2.8	11.9 ± 3.7	38.1 ± 9.3
P3Mix-E	2.2 ± 2.8	46.3 ± 8.1	9.5 ± 10.3	5.6 ± 1.6	2.0 ± 2.4	7.2 ± 6.2	9.7 ± 7.3	10.9 ± 3.1	39.2 ± 8.2
PUET (KM2)	11.0 ± 10.5	62.9 ± 10.3	35.0 ± 25.8	3.4 ± 2.0	46.0 ± 30.3	44.0 ± 26.4	65.0 ± 9.9	5.2 ± 2.1	45.6 ± 9.9
ScalePU (KM2)	0.0 ± 0.0	29.5 ± 15.2	3.4 ± 5.9	2.0 ± 1.4	4.6 ± 3.9	19.7 ± 24.0	20.4 ± 10.7	0.0 ± 0.0	19.6 ± 4.9
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	22.9 ± 32.4	0.0 ± 0.0	37.6 ± 39.7	28.8 ± 29.4	12.1 ± 13.7	0.0 ± 0.0	0.0 ± 0.0
Dist-PU (KM2)	15.4 ± 13.1	77.6 ± 9.2	24.6 ± 32.9	3.7 ± 1.8	49.9 ± 30.9	56.0 ± 24.9	87.3 ± 4.1	0.0 ± 0.0	32.7 ± 12.8
PULDA (KM2)	5.9 ± 8.1	**89.7** ± 6.3	39.7 ± 24.5	3.5 ± 1.7	44.2 ± 36.7	11.5 ± 10.9	7.9 ± 7.1	2.5 ± 2.8	40.4 ± 8.7
UREA (KM2)	20.0 ± 8.4	62.5 ± 18.0	59.4 ± 17.2	24.1 ± 16.6	42.5 ± 17.3	60.1 ± 12.7	20.3 ± 10.2	2.1 ± 1.8	26.6 ± 9.4
AREA (KM2)	3.9 ± 5.8	23.2 ± 8.6	14.2 ± 8.9	1.7 ± 1.2	2.6 ± 2.3	6.6 ± 7.6	14.6 ± 6.8	2.8 ± 1.5	27.1 ± 9.8
PUET (real)	4.7 ± 4.8	88.7 ± 5.2	69.5 ± 7.7	45.5 ± 10.6	62.8 ± 4.8	85.8 ± 3.4	74.1 ± 2.7	10.9 ± 2.2	49.4 ± 6.4
ScalePU (real)	0.6 ± 1.9	30.8 ± 13.9	10.8 ± 8.4	2.3 ± 1.7	5.1 ± 4.1	43.8 ± 23.0	18.4 ± 11.8	0.2 ± 0.5	18.7 ± 4.2
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	62.9 ± 13.0	0.0 ± 0.0	66.3 ± 4.3	87.8 ± 4.9	29.8 ± 8.7	0.0 ± 0.0	0.0 ± 0.0
Dist-PU (real)	15.3 ± 8.9	88.0 ± 5.6	73.8 ± 9.9	24.6 ± 9.5	72.2 ± 6.9	79.4 ± 7.8	94.7 ± 1.1	3.0 ± 3.0	16.4 ± 2.4
PULDA (real)	6.6 ± 8.0	90.9 ± 5.5	68.8 ± 0.0	10.8 ± 8.9	78.9 ± 0.0	23.6 ± 13.5	6.3 ± 5.0	16.2 ± 3.5	42.9 ± 12.6
UREA (real)	18.2 ± 7.9	80.8 ± 5.2	64.7 ± 14.3	53.6 ± 4.3	57.4 ± 7.2	74.8 ± 7.0	18.0 ± 2.9	13.8 ± 3.8	20.7 ± 4.2
AREA (real)	3.5 ± 4.9	31.7 ± 12.2	14.5 ± 8.7	2.4 ± 1.4	3.6 ± 2.6	6.7 ± 5.5	9.6 ± 6.1	12.0 ± 3.0	32.1 ± 7.5

	Gesture Phase Seg. Proc.	first-order theorem prov.	mfeat-karhunen	mfeat-fourier	MiceProtein	satimage	steel-plates fault	Average Pos.-F1 (%)	Average Rank	Wins
DRIP-MPU (Ours)	28.6 ± 1.9	**21.1** ± 2.6	**88.9** ± 3.0	**75.3** ± 2.3	**98.2** ± 1.1	**74.8** ± 2.5	**65.4** ± 2.9	**59.7** ± 1.0	**1.640**	**18**
BBE-PU	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	17.4 ± 6.1	0.0 ± 0.0	2.0 ± 3.8	11.2 ± 1.5	10.980	1
LaGAM-PU	**29.8** ± 2.4	20.1 ± 3.9	78.3 ± 3.4	55.0 ± 3.7	70.2 ± 3.3	60.6 ± 5.3	47.3 ± 3.3	43.7 ± 1.8	3.760	2
CGenPU	17.5 ± 3.0	13.5 ± 3.4	23.2 ± 7.5	9.6 ± 4.4	18.3 ± 7.5	17.3 ± 6.8	45.6 ± 4.9	39.2 ± 1.7	4.080	0
VPU	0.0 ± 0.1	1.4 ± 0.9	1.0 ± 0.6	0.5 ± 0.5	0.2 ± 0.6	3.7 ± 3.3	13.0 ± 5.1	6.6 ± 1.5	13.040	0
PAN	0.0 ± 0.0	0.4 ± 1.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	5.2 ± 6.7	14.0 ± 4.2	11.400	0
HolisticPU	10.6 ± 2.9	12.7 ± 3.0	1.6 ± 0.6	1.9 ± 1.0	0.5 ± 0.5	8.4 ± 4.5	4.9 ± 1.3	20.7 ± 0.9	8.120	2
P3Mix-C	5.1 ± 1.6	12.1 ± 2.2	6.9 ± 1.9	5.3 ± 0.8	7.5 ± 2.3	13.7 ± 2.8	28.2 ± 5.1	11.7 ± 1.0	10.000	0
P3Mix-E	4.4 ± 1.8	13.1 ± 2.2	5.7 ± 1.6	4.9 ± 0.9	5.9 ± 2.1	11.1 ± 2.2	28.8 ± 6.2	11.0 ± 1.1	10.480	0
PUET (KM2)	10.7 ± 6.9	11.4 ± 2.5	29.1 ± 5.7	27.3 ± 4.6	65.5 ± 6.6	66.7 ± 5.4	40.1 ± 7.5	32.2 ± 4.2	5.960	0
ScalePU (KM2)	0.4 ± 0.2	4.5 ± 2.3	0.3 ± 0.3	0.3 ± 0.3	2.6 ± 1.2	8.2 ± 4.6	22.3 ± 5.8	6.8 ± 1.6	12.780	0
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	13.8 ± 9.2	0.0 ± 0.0	0.0 ± 0.0	11.0 ± 4.5	12.200	0
Dist-PU (KM2)	8.2 ± 5.6	11.3 ± 4.0	73.8 ± 3.3	50.7 ± 4.5	76.8 ± 5.0	54.6 ± 4.0	33.2 ± 7.2	35.8 ± 3.6	6.100	1
PULDA (KM2)	7.9 ± 6.0	16.0 ± 4.5	8.8 ± 1.8	8.3 ± 2.5	65.5 ± 6.2	45.5 ± 6.2	50.3 ± 5.9	26.7 ± 3.2	7.180	1
UREA (KM2)	9.1 ± 1.6	10.3 ± 3.5	13.2 ± 4.4	14.0 ± 5.2	11.4 ± 5.2	48.4 ± 10.1	20.4 ± 7.0	29.1 ± 3.0	6.600	0
AREA (KM2)	2.2 ± 1.9	6.7 ± 2.6	7.3 ± 1.5	4.7 ± 1.3	0.1 ± 0.3	21.8 ± 3.8	26.9 ± 4.7	8.4 ± 1.4	11.680	0
PUET (real)	7.2 ± 2.2	14.0 ± 2.0	21.9 ± 3.9	19.3 ± 4.1	46.5 ± 4.7	70.2 ± 3.6	55.4 ± 6.9	47.3 ± 0.7	–	–
ScalePU (real)	0.3 ± 0.2	4.7 ± 2.0	0.3 ± 0.3	0.1 ± 0.2	1.2 ± 1.2	10.4 ± 3.6	26.6 ± 5.9	9.6 ± 1.7	–	–
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	21.9 ± 2.1	–	–
Dist-PU (real)	2.3 ± 0.9	15.3 ± 2.8	72.8 ± 2.8	46.1 ± 3.4	74.7 ± 4.1	60.1 ± 3.5	52.2 ± 3.2	49.8 ± 1.9	–	–
PULDA (real)	1.7 ± 0.8	18.6 ± 4.1	6.3 ± 1.8	6.1 ± 1.6	41.3 ± 4.8	52.3 ± 4.8	53.4 ± 6.4	36.4 ± 1.2	–	–
UREA (real)	10.5 ± 0.9	10.2 ± 2.1	17.1 ± 2.5	16.5 ± 6.4	12.4 ± 7.6	54.1 ± 7.4	20.6 ± 6.6	38.8 ± 1.7	–	–
AREA (real)	2.5 ± 0.9	8.3 ± 2.3	7.0 ± 1.3	5.1 ± 2.1	0.3 ± 0.5	21.4 ± 3.1	27.7 ± 6.7	9.7 ± 1.4	–	–

Table 9: AUC (%) results on all 25 datasets in the 10-shot setting.

Method	Australian	blood-transfusion	climate-model sim.-crashes	credit-approval	credit-g	cylinder-bands	diabetes	dresses-sales	kc2
DRIP-MPU (Ours)	**84.8** ± 8.3	64.6 ± 11.2	51.2 ± 13.2	**85.2** ± 5.4	**67.6** ± 4.5	62.6 ± 9.4	**73.6** ± 4.3	50.6 ± 5.9	68.6 ± 13.4
BBE-PU	73.3 ± 10.4	**68.4** ± 7.9	56.2 ± 11.7	79.1 ± 4.7	60.0 ± 4.4	57.2 ± 6.4	70.2 ± 5.4	53.2 ± 4.5	70.6 ± 3.0
LaGAM-PU	66.4 ± 16.8	63.2 ± 11.5	56.3 ± 15.5	72.7 ± 15.3	56.8 ± 4.2	54.2 ± 9.7	66.0 ± 8.1	52.8 ± 4.6	63.8 ± 11.4
CGenPU	71.3 ± 12.3	56.6 ± 13.2	50.5 ± 10.9	69.8 ± 10.6	57.7 ± 6.5	63.9 ± 8.3	65.4 ± 8.2	50.1 ± 4.6	53.2 ± 14.4
VPU	58.8 ± 9.1	63.2 ± 8.3	**57.5** ± 9.2	70.1 ± 7.1	54.6 ± 4.5	55.0 ± 5.0	60.8 ± 4.9	**54.8** ± 6.6	55.8 ± 9.2
PAN	73.8 ± 17.7	65.5 ± 9.3	51.6 ± 13.4	77.8 ± 7.6	49.6 ± 5.1	52.1 ± 12.7	68.3 ± 7.5	53.0 ± 5.5	72.7 ± 4.2
HolisticPU	29.0 ± 13.6	48.2 ± 13.7	43.6 ± 11.5	23.8 ± 7.0	40.5 ± 5.2	46.4 ± 9.0	32.9 ± 7.6	49.3 ± 5.1	35.6 ± 6.7
P3Mix-C	60.9 ± 8.2	60.0 ± 4.7	55.0 ± 10.6	74.1 ± 7.0	59.5 ± 7.6	56.4 ± 6.5	67.4 ± 5.9	54.4 ± 5.6	63.6 ± 8.3
P3Mix-E	62.4 ± 11.7	62.2 ± 6.3	56.2 ± 11.0	74.1 ± 7.9	57.9 ± 4.7	53.3 ± 7.4	67.8 ± 6.2	53.3 ± 5.4	55.6 ± 9.8
PUET (KM2)	74.9 ± 11.7	57.5 ± 7.7	51.2 ± 12.9	78.0 ± 6.1	62.8 ± 7.8	**65.6** ± 9.9	67.9 ± 3.7	51.9 ± 4.8	70.8 ± 4.7
ScalePU (KM2)	70.1 ± 10.7	62.8 ± 12.4	51.6 ± 11.9	76.3 ± 8.3	59.3 ± 5.7	61.7 ± 5.5	69.6 ± 6.9	53.0 ± 5.1	68.1 ± 6.8
nnPU (KM2)	57.6 ± 10.6	53.4 ± 5.3	56.6 ± 9.2	72.9 ± 8.3	51.5 ± 7.5	46.4 ± 11.7	51.6 ± 8.2	52.7 ± 6.5	68.1 ± 9.4
Dist-PU (KM2)	70.9 ± 18.2	62.5 ± 8.3	45.8 ± 9.0	80.7 ± 9.4	58.1 ± 8.7	49.7 ± 14.9	67.0 ± 11.7	51.7 ± 7.1	**73.6** ± 1.0
PULDA (KM2)	73.5 ± 10.3	66.1 ± 7.2	45.7 ± 6.6	78.2 ± 7.9	58.7 ± 3.3	59.4 ± 7.7	68.8 ± 5.1	51.9 ± 5.8	65.7 ± 8.7
UREA (KM2)	58.8 ± 6.3	60.9 ± 7.7	53.2 ± 11.7	62.1 ± 4.4	57.8 ± 3.8	58.8 ± 3.9	55.9 ± 4.0	51.5 ± 3.6	57.9 ± 8.7
AREA (KM2)	61.1 ± 11.7	59.9 ± 7.6	51.1 ± 11.3	68.6 ± 9.1	57.2 ± 5.0	59.5 ± 5.3	59.9 ± 7.8	51.5 ± 4.8	51.4 ± 9.3
PUET (real)	82.3 ± 8.7	61.5 ± 6.7	60.5 ± 11.7	82.0 ± 6.6	65.0 ± 4.4	68.6 ± 8.7	72.6 ± 4.5	52.7 ± 7.2	73.2 ± 4.7
ScalePU (real)	69.7 ± 10.8	62.6 ± 12.2	51.6 ± 11.7	76.2 ± 8.4	59.3 ± 5.5	61.6 ± 5.6	69.5 ± 6.9	53.1 ± 5.1	68.2 ± 6.8
nnPU (real)	78.6 ± 12.7	62.0 ± 6.8	53.9 ± 11.0	80.6 ± 6.8	53.5 ± 3.8	61.3 ± 11.5	73.9 ± 5.7	53.3 ± 4.8	73.3 ± 2.6
Dist-PU (real)	76.9 ± 13.3	69.8 ± 9.1	45.4 ± 8.9	78.2 ± 8.5	60.0 ± 5.5	60.5 ± 11.0	71.9 ± 5.7	53.1 ± 5.3	73.1 ± 1.8
PULDA (real)	79.4 ± 8.3	66.8 ± 5.2	45.7 ± 5.8	80.3 ± 9.2	58.4 ± 3.8	69.3 ± 5.5	69.3 ± 5.5	52.9 ± 5.8	65.8 ± 8.5
UREA (real)	67.4 ± 8.8	60.7 ± 8.7	56.3 ± 11.0	70.3 ± 7.5	56.0 ± 3.6	60.1 ± 4.8	63.3 ± 5.2	49.9 ± 4.4	57.0 ± 7.5
AREA (real)	64.2 ± 15.3	61.6 ± 7.9	53.3 ± 11.7	71.0 ± 9.6	58.1 ± 4.9	60.6 ± 6.7	61.3 ± 7.0	51.8 ± 4.9	52.0 ± 8.5

	pc3	segment	sonar	spambase	tic-tac-toe	wdbc	analcatdata_author	analcatdata_dmft	car
DRIP-MPU (Ours)	72.0 ± 8.8	**99.2** ± 0.9	78.3 ± 8.6	**88.1** ± 4.6	54.1 ± 7.0	**98.7** ± 1.3	**99.9** ± 0.1	51.5 ± 2.8	89.5 ± 2.7
BBE-PU	74.3 ± 4.2	96.5 ± 2.4	78.0 ± 8.4	61.8 ± 4.9	55.0 ± 8.1	93.7 ± 3.4	93.9 ± 3.2	51.6 ± 2.4	85.5 ± 3.6
LaGAM-PU	68.0 ± 9.3	91.2 ± 5.2	70.4 ± 13.1	50.2 ± 4.0	54.4 ± 5.5	80.1 ± 10.1	96.9 ± 0.7	**52.9** ± 2.9	83.4 ± 2.7
CGenPU	57.8 ± 14.0	82.7 ± 8.9	63.9 ± 14.7	57.1 ± 12.9	53.7 ± 4.5	85.9 ± 13.2	72.5 ± 9.0	51.2 ± 2.2	77.2 ± 3.7
VPU	61.8 ± 6.6	94.2 ± 2.7	70.5 ± 10.1	63.5 ± 2.3	53.8 ± 6.4	77.4 ± 4.2	74.3 ± 1.9	51.9 ± 2.6	85.9 ± 2.7
PAN	74.2 ± 10.4	96.5 ± 4.1	54.4 ± 13.0	59.4 ± 6.6	53.0 ± 5.4	96.0 ± 5.3	71.0 ± 3.3	51.6 ± 4.0	58.1 ± 7.4
HolisticPU	29.9 ± 7.7	33.4 ± 6.7	28.9 ± 12.3	62.9 ± 3.6	46.2 ± 7.9	5.6 ± 4.2	7.1 ± 2.1	50.0 ± 2.8	23.4 ± 4.2
P3Mix-C	67.6 ± 7.8	96.2 ± 2.0	66.5 ± 12.5	68.5 ± 4.9	54.7 ± 7.2	82.9 ± 4.5	72.9 ± 2.6	51.8 ± 1.8	86.9 ± 3.2
P3Mix-E	60.7 ± 7.3	97.6 ± 2.5	67.2 ± 10.7	59.0 ± 3.9	55.5 ± 7.4	82.9 ± 8.4	69.3 ± 3.0	51.3 ± 2.0	82.1 ± 3.8
PUET (KM2)	70.5 ± 7.1	98.9 ± 1.9	73.2 ± 9.3	71.2 ± 7.5	55.4 ± 6.2	96.0 ± 2.0	98.5 ± 0.5	50.5 ± 1.7	83.4 ± 3.8
ScalePU (KM2)	64.9 ± 6.7	95.6 ± 3.1	**78.8** ± 7.4	60.7 ± 2.7	55.8 ± 7.2	93.9 ± 6.3	86.2 ± 3.1	51.8 ± 2.8	78.0 ± 4.6
nnPU (KM2)	67.1 ± 6.6	76.3 ± 3.6	72.0 ± 5.1	50.7 ± 1.2	53.4 ± 5.9	96.0 ± 5.3	90.5 ± 4.8	49.5 ± 2.4	65.6 ± 5.7
Dist-PU (KM2)	**75.8** ± 3.5	96.7 ± 2.5	70.7 ± 9.8	58.5 ± 4.4	55.1 ± 7.3	97.0 ± 2.5	98.5 ± 0.8	50.7 ± 2.5	81.2 ± 11.5
PULDA (KM2)	65.5 ± 11.9	99.1 ± 0.6	68.2 ± 15.4	54.9 ± 6.3	**55.9** ± 7.2	88.1 ± 5.2	91.0 ± 1.7	51.8 ± 2.8	88.9 ± 3.9
UREA (KM2)	55.1 ± 3.9	89.7 ± 4.7	66.1 ± 9.8	56.2 ± 4.9	51.6 ± 5.9	71.8 ± 6.4	60.1 ± 4.7	50.7 ± 2.1	77.4 ± 7.4
AREA (KM2)	57.2 ± 8.5	93.6 ± 4.2	69.5 ± 6.7	51.9 ± 7.1	53.3 ± 6.5	82.0 ± 6.1	69.5 ± 6.0	51.2 ± 3.1	**90.8** ± 2.0
PUET (real)	70.8 ± 5.3	99.3 ± 1.2	78.4 ± 7.1	85.2 ± 4.5	55.3 ± 6.6	98.5 ± 0.5	99.0 ± 0.5	51.1 ± 2.2	82.0 ± 2.2
ScalePU (real)	65.0 ± 6.6	95.7 ± 2.8	78.9 ± 7.5	60.4 ± 3.2	55.8 ± 7.2	93.7 ± 6.6	86.1 ± 3.2	51.8 ± 2.6	78.0 ± 4.1
nnPU (real)	65.2 ± 4.5	81.4 ± 2.9	78.0 ± 7.0	54.3 ± 2.1	55.7 ± 7.2	98.8 ± 0.8	89.8 ± 2.9	49.8 ± 2.2	55.2 ± 3.5
Dist-PU (real)	75.7 ± 3.1	96.9 ± 2.4	81.6 ± 8.4	54.2 ± 8.3	55.3 ± 6.7	96.3 ± 4.7	99.3 ± 0.5	51.3 ± 3.6	73.8 ± 9.5
PULDA (real)	65.8 ± 12.4	99.1 ± 0.6	49.1 ± 2.9	43.9 ± 3.6	53.5 ± 4.9	91.2 ± 3.9	91.9 ± 1.8	52.2 ± 2.9	90.8 ± 4.8
UREA (real)	55.3 ± 3.8	88.1 ± 3.5	69.4 ± 10.7	63.9 ± 3.0	52.7 ± 6.1	80.4 ± 4.3	59.2 ± 2.8	49.7 ± 3.7	64.1 ± 2.4
AREA (real)	54.8 ± 10.1	95.4 ± 2.7	68.9 ± 6.6	58.7 ± 4.8	54.3 ± 6.8	84.1 ± 7.8	70.5 ± 5.9	51.0 ± 2.8	90.2 ± 2.2

	Gesture Phase Seg. Proc.	first-order theorem prov.	mfeat-karhunen	mfeat-fourier	MiceProtein	satimage	steel-plates fault	Average AUC (%)	Average Rank	Wins
DRIP-MPU (Ours)	57.1 ± 2.2	58.3 ± 2.6	**98.2** ± 1.2	**96.5** ± 1.0	**99.2** ± 0.8	**96.2** ± 0.6	87.4 ± 1.3	**77.3** ± 1.3	**3.280**	**12**
BBE-PU	57.2 ± 2.9	54.1 ± 1.9	95.2 ± 0.6	88.6 ± 1.3	94.0 ± 1.8	77.2 ± 7.4	85.9 ± 2.4	73.2 ± 1.3	5.200	1
LaGAM-PU	58.9 ± 2.0	**59.5** ± 2.2	94.9 ± 1.0	89.6 ± 1.2	91.8 ± 1.9	91.7 ± 0.8	84.5 ± 1.2	70.8 ± 2.1	7.280	2
CGenPU	53.0 ± 2.4	53.5 ± 2.7	71.9 ± 5.3	60.6 ± 4.7	67.8 ± 4.8	64.3 ± 6.5	80.0 ± 1.9	63.7 ± 2.0	11.880	0
VPU	56.4 ± 2.3	58.1 ± 1.4	86.0 ± 1.3	80.0 ± 1.4	87.6 ± 0.9	84.4 ± 1.1	85.9 ± 1.4	68.1 ± 1.1	7.960	2
PAN	**60.7** ± 5.7	54.2 ± 3.1	80.1 ± 1.6	62.5 ± 2.5	82.4 ± 1.2	71.6 ± 5.0	81.5 ± 2.2	66.9 ± 1.8	8.480	1
HolisticPU	49.0 ± 1.8	49.0 ± 2.7	13.4 ± 1.1	16.7 ± 1.0	14.4 ± 2.6	33.5 ± 10.9	26.3 ± 4.2	33.6 ± 1.5	15.440	0
P3Mix-C	56.8 ± 4.2	56.5 ± 1.5	80.0 ± 1.8	75.3 ± 1.6	79.7 ± 2.5	85.4 ± 1.0	83.7 ± 2.3	68.7 ± 1.6	7.880	0
P3Mix-E	55.4 ± 4.0	56.4 ± 1.5	74.8 ± 2.6	70.7 ± 1.9	83.7 ± 1.6	82.7 ± 1.6	83.3 ± 1.9	67.0 ± 1.4	9.000	0
PUET (KM2)	60.6 ± 5.3	57.7 ± 1.6	94.6 ± 0.6	91.3 ± 0.7	96.4 ± 1.5	94.9 ± 0.6	86.9 ± 1.9	74.4 ± 1.5	4.840	1
ScalePU (KM2)	50.7 ± 1.5	55.0 ± 1.6	83.5 ± 1.7	79.3 ± 1.7	89.9 ± 1.4	82.0 ± 1.5	84.0 ± 2.0	70.5 ± 1.7	7.120	1
nnPU (KM2)	54.7 ± 2.7	52.0 ± 1.7	89.6 ± 1.3	82.9 ± 8.3	73.7 ± 3.0	63.1 ± 3.6	75.1 ± 3.6	64.9 ± 0.9	11.120	0
Dist-PU (KM2)	56.3 ± 4.7	57.4 ± 2.2	97.0 ± 0.5	91.8 ± 1.4	96.0 ± 1.5	94.0 ± 1.3	86.9 ± 3.2	72.9 ± 1.3	6.280	2
PULDA (KM2)	55.4 ± 4.3	57.8 ± 3.1	94.6 ± 1.3	87.5 ± 1.7	96.9 ± 1.6	90.5 ± 1.1	**88.6** ± 1.1	72.1 ± 1.0	5.880	2
UREA (KM2)	52.2 ± 1.8	53.9 ± 1.8	55.8 ± 3.0	57.7 ± 4.0	55.2 ± 2.9	75.7 ± 6.6	58.2 ± 5.4	60.2 ± 1.7	12.720	0
AREA (KM2)	43.8 ± 1.9	54.9 ± 2.2	64.1 ± 2.0	64.5 ± 2.5	62.5 ± 1.5	83.0 ± 3.2	74.5 ± 3.3	63.5 ± 1.3	11.640	1
PUET (real)	65.6 ± 1.9	58.6 ± 2.0	94.9 ± 1.2	90.8 ± 1.2	97.5 ± 0.9	95.2 ± 0.4	87.8 ± 1.8	77.1 ± 1.0	–	–
ScalePU (real)	50.6 ± 1.5	54.8 ± 1.4	83.6 ± 1.7	79.3 ± 1.6	89.9 ± 1.3	81.4 ± 1.2	83.8 ± 1.8	70.4 ± 1.7	–	–
nnPU (real)	54.5 ± 1.3	52.0 ± 1.2	90.6 ± 0.8	86.2 ± 1.4	87.2 ± 1.3	64.2 ± 2.5	76.1 ± 5.1	69.2 ± 1.0	–	–
Dist-PU (real)	50.3 ± 1.9	58.6 ± 2.5	97.3 ± 0.4	91.7 ± 1.2	97.0 ± 1.0	94.4 ± 0.9	88.4 ± 2.2	74.1 ± 1.3	–	–
PULDA (real)	48.7 ± 2.7	58.5 ± 2.7	94.5 ± 0.8	86.6 ± 1.6	98.1 ± 0.5	89.8 ± 0.9	88.8 ± 1.1	71.6 ± 0.6	–	–
UREA (real)	51.1 ± 2.3	52.6 ± 1.7	57.6 ± 2.5	60.7 ± 3.2	56.9 ± 4.1	78.3 ± 4.0	58.6 ± 2.2	61.6 ± 1.5	–	–
AREA (real)	42.7 ± 0.8	55.7 ± 1.9	63.6 ± 2.0	64.8 ± 2.4	61.8 ± 2.4	82.3 ± 3.2	74.4 ± 2.8	64.3 ± 1.5	–	–

Table 10: Pos.-F1 (%) results on all 25 datasets in the 15-shot setting.

Method	Australian	blood-transfusion	climate-model sim.-crashes	credit-approval	credit-g	cylinder-bands	diabetes	dresses-sales	kc2
DRIP-MPU (Ours)	**76.5** ± 3.6	45.0 ± 5.8	67.2 ± 7.0	**78.5** ± 1.5	**52.2** ± 6.6	**58.9** ± 10.3	**60.9** ± 5.2	47.4 ± 6.4	42.7 ± 6.9
BBE-PU	34.7 ± 14.7	**49.7** ± 6.2	2.2 ± 6.3	31.8 ± 13.5	10.9 ± 9.3	5.2 ± 6.0	34.6 ± 13.1	11.6 ± 9.1	33.2 ± 15.4
LaGAM-PU	46.5 ± 12.7	45.9 ± 4.1	46.0 ± 4.8	58.5 ± 11.4	33.2 ± 5.2	37.6 ± 7.5	41.9 ± 7.2	31.1 ± 5.7	39.8 ± 6.7
CGenPU	63.5 ± 9.8	40.8 ± 3.1	66.9 ± 6.2	61.0 ± 3.7	47.0 ± 5.8	54.6 ± 7.3	53.0 ± 7.3	**47.8** ± 6.0	35.7 ± 5.7
VPU	3.2 ± 3.3	28.3 ± 17.7	0.8 ± 1.7	5.0 ± 3.5	1.3 ± 1.7	2.5 ± 2.9	5.2 ± 4.1	1.4 ± 2.2	13.4 ± 12.5
PAN	70.9 ± 7.1	41.1 ± 5.7	**91.3** ± 9.3	59.7 ± 22.3	0.0 ± 0.0	52.1 ± 13.6	51.2 ± 18.7	32.7 ± 28.5	27.1 ± 19.5
HolisticPU	28.2 ± 6.1	16.7 ± 6.6	65.8 ± 6.6	31.2 ± 5.0	29.8 ± 5.5	45.1 ± 8.4	31.3 ± 6.4	45.5 ± 5.5	15.2 ± 5.5
P3Mix-C	6.1 ± 6.0	33.1 ± 8.6	1.4 ± 1.9	10.5 ± 6.7	3.8 ± 3.5	7.7 ± 5.7	10.3 ± 3.7	5.2 ± 5.6	16.6 ± 9.2
P3Mix-E	5.4 ± 6.8	26.7 ± 10.3	0.8 ± 1.4	7.6 ± 6.1	3.8 ± 4.5	6.8 ± 6.2	6.4 ± 3.6	4.0 ± 4.3	11.9 ± 6.4
PUET (KM2)	27.0 ± 13.7	22.1 ± 13.1	71.9 ± 38.6	25.9 ± 16.2	20.9 ± 21.0	32.8 ± 23.8	19.6 ± 16.6	23.8 ± 29.8	39.0 ± 12.5
ScalePU (KM2)	9.6 ± 8.0	0.0 ± 0.0	12.1 ± 29.5	12.8 ± 11.2	1.6 ± 2.7	3.3 ± 4.2	13.4 ± 15.7	3.0 ± 4.3	9.5 ± 20.3
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	70.9 ± 39.6	6.1 ± 19.4	13.9 ± 22.4	24.0 ± 28.2	11.5 ± 24.2	23.5 ± 30.3	35.6 ± 18.8
Dist-PU (KM2)	37.2 ± 23.4	4.7 ± 14.8	77.8 ± 29.1	19.4 ± 24.1	24.4 ± 21.9	30.9 ± 27.9	21.4 ± 23.6	25.3 ± 29.6	**43.8** ± 6.4
PULDA (KM2)	31.3 ± 13.3	19.2 ± 23.7	76.5 ± 40.3	31.6 ± 16.4	18.1 ± 20.3	32.3 ± 24.9	27.1 ± 16.1	25.0 ± 29.5	40.0 ± 8.4
UREA (KM2)	43.0 ± 13.1	14.3 ± 18.4	61.3 ± 24.0	45.3 ± 13.6	34.2 ± 13.2	41.5 ± 15.4	36.6 ± 12.3	18.5 ± 21.3	43.2 ± 7.1
AREA (KM2)	5.3 ± 4.0	7.4 ± 10.0	3.7 ± 1.9	8.4 ± 3.8	2.9 ± 3.2	8.8 ± 5.7	6.5 ± 6.0	0.9 ± 1.9	12.1 ± 5.6
PUET (real)	72.5 ± 9.2	28.7 ± 10.6	95.6 ± 0.2	72.7 ± 5.6	29.1 ± 8.0	54.3 ± 11.8	40.9 ± 6.2	33.9 ± 10.8	38.4 ± 10.8
ScalePU (real)	14.1 ± 6.4	0.0 ± 0.0	12.3 ± 29.5	18.4 ± 9.4	2.6 ± 2.1	3.3 ± 4.9	29.6 ± 15.8	3.5 ± 3.4	0.0 ± 0.0
nnPU (real)	46.7 ± 32.3	0.0 ± 0.0	95.6 ± 0.2	66.1 ± 8.8	0.0 ± 0.0	48.2 ± 9.9	37.8 ± 18.7	30.0 ± 6.5	26.7 ± 19.3
Dist-PU (real)	70.5 ± 11.0	12.5 ± 20.2	95.7 ± 0.0	69.3 ± 5.9	36.5 ± 6.5	52.4 ± 10.0	49.4 ± 5.6	38.4 ± 5.3	45.3 ± 6.2
PULDA (real)	70.7 ± 9.6	48.5 ± 7.0	95.7 ± 0.0	68.0 ± 13.8	12.1 ± 7.7	59.7 ± 0.0	44.0 ± 10.9	33.1 ± 22.8	37.7 ± 11.2
UREA (real)	64.3 ± 8.5	32.3 ± 14.3	77.3 ± 6.8	64.8 ± 7.1	35.0 ± 9.2	50.0 ± 6.4	50.9 ± 5.9	32.2 ± 5.9	43.3 ± 7.0
AREA (real)	7.4 ± 4.5	22.7 ± 12.6	4.9 ± 2.1	7.4 ± 3.5	2.6 ± 2.5	6.1 ± 4.3	7.7 ± 6.5	0.9 ± 1.9	9.3 ± 4.7

	pc3	segment	sonar	spambase	tic-tac-toe	wdbc	analcatdata_author	analcatdata_dmft	car
DRIP-MPU (Ours)	**29.4** ± 5.9	50.0 ± 17.7	**74.7** ± 6.9	**77.8** ± 4.0	61.0 ± 5.5	**90.8** ± 4.5	**97.8** ± 1.4	16.8 ± 2.3	**55.1** ± 4.2
BBE-PU	17.6 ± 7.5	50.4 ± 14.0	9.3 ± 13.4	1.4 ± 1.6	21.2 ± 15.6	54.8 ± 16.2	23.3 ± 8.4	15.8 ± 3.5	43.8 ± 11.0
LaGAM-PU	27.3 ± 8.4	78.2 ± 5.2	58.6 ± 9.2	39.7 ± 7.4	40.1 ± 7.6	71.7 ± 7.7	73.5 ± 3.5	**20.4** ± 3.0	27.0 ± 4.0
CGenPU	19.7 ± 5.3	42.9 ± 17.5	69.3 ± 2.0	49.9 ± 8.8	**65.3** ± 5.1	73.4 ± 8.7	52.1 ± 12.2	17.3 ± 4.0	32.6 ± 5.0
VPU	1.2 ± 2.5	64.8 ± 12.6	9.8 ± 10.2	2.1 ± 1.4	1.1 ± 1.5	3.1 ± 5.9	7.5 ± 4.1	12.4 ± 2.5	41.5 ± 7.5
PAN	0.0 ± 0.0	4.7 ± 14.7	45.7 ± 32.4	0.0 ± 0.0	54.4 ± 28.1	76.4 ± 15.9	0.0 ± 0.0	18.4 ± 5.0	5.3 ± 8.5
HolisticPU	7.3 ± 2.2	3.6 ± 1.3	36.3 ± 8.8	4.0 ± 1.5	56.5 ± 4.9	9.8 ± 3.8	3.9 ± 0.9	11.9 ± 2.4	6.1 ± 1.5
P3Mix-C	6.3 ± 5.4	52.9 ± 8.1	20.9 ± 10.4	7.5 ± 1.8	4.9 ± 3.8	10.5 ± 6.9	16.8 ± 3.9	14.4 ± 2.1	48.2 ± 5.5
P3Mix-E	5.5 ± 4.5	56.0 ± 9.8	18.7 ± 5.9	6.8 ± 1.5	2.7 ± 2.2	7.5 ± 7.6	18.8 ± 7.6	13.2 ± 2.5	45.8 ± 7.2
PUET (KM2)	17.2 ± 9.5	65.7 ± 13.8	35.1 ± 21.3	7.4 ± 3.4	35.2 ± 21.3	73.7 ± 21.4	76.3 ± 7.9	7.1 ± 2.5	51.9 ± 8.8
ScalePU (KM2)	3.4 ± 4.1	44.5 ± 20.8	19.0 ± 24.3	2.9 ± 1.2	9.1 ± 7.3	53.5 ± 22.4	31.4 ± 13.3	0.0 ± 0.0	33.5 ± 11.1
nnPU (KM2)	2.1 ± 6.7	0.0 ± 0.0	14.6 ± 28.7	0.0 ± 0.0	16.5 ± 30.8	71.9 ± 23.2	28.6 ± 18.0	0.0 ± 0.0	0.0 ± 0.0
Dist-PU (KM2)	16.5 ± 10.4	80.5 ± 12.2	18.2 ± 28.9	9.8 ± 5.1	38.7 ± 22.5	78.6 ± 9.7	91.2 ± 3.2	0.2 ± 0.8	39.6 ± 17.6
PULDA (KM2)	12.1 ± 9.1	**92.3** ± 6.0	54.4 ± 17.2	7.3 ± 2.2	30.6 ± 27.5	41.0 ± 20.0	14.4 ± 6.1	5.8 ± 3.6	48.6 ± 10.8
UREA (KM2)	23.1 ± 4.1	68.4 ± 11.7	59.8 ± 13.3	31.3 ± 14.7	37.6 ± 8.1	71.9 ± 11.9	25.0 ± 12.8	4.7 ± 3.2	24.9 ± 8.3
AREA (KM2)	5.0 ± 5.4	37.7 ± 13.7	23.2 ± 9.6	2.8 ± 0.7	3.4 ± 2.6	9.7 ± 5.6	16.9 ± 4.1	5.5 ± 3.3	42.2 ± 12.2
PUET (real)	8.5 ± 5.7	92.6 ± 4.1	73.5 ± 5.7	57.3 ± 6.7	67.2 ± 3.5	86.5 ± 8.4	81.8 ± 4.0	11.9 ± 3.3	57.2 ± 3.9
ScalePU (real)	4.6 ± 4.6	53.3 ± 14.3	34.7 ± 15.1	3.0 ± 1.3	14.0 ± 7.3	60.5 ± 18.9	25.9 ± 19.3	0.0 ± 0.0	30.0 ± 8.2
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	60.1 ± 10.1	0.0 ± 0.0	67.3 ± 3.3	91.8 ± 4.1	40.1 ± 10.4	0.0 ± 0.0	0.0 ± 0.0
Dist-PU (real)	14.5 ± 7.0	91.3 ± 5.7	75.2 ± 7.6	35.3 ± 9.3	71.3 ± 5.5	86.0 ± 3.8	95.8 ± 1.0	1.1 ± 1.6	18.6 ± 3.0
PULDA (real)	8.4 ± 6.6	95.1 ± 3.5	68.8 ± 0.0	20.3 ± 6.3	78.9 ± 0.0	50.6 ± 18.8	14.5 ± 5.0	15.6 ± 2.7	54.5 ± 7.6
UREA (real)	20.9 ± 5.4	86.0 ± 4.7	70.5 ± 10.8	62.1 ± 3.4	61.6 ± 5.3	81.4 ± 6.7	21.2 ± 3.5	16.3 ± 2.5	21.4 ± 2.5
AREA (real)	5.0 ± 5.6	45.2 ± 12.8	21.4 ± 8.0	3.2 ± 1.3	5.9 ± 3.7	8.0 ± 6.2	19.2 ± 3.4	14.1 ± 2.7	37.9 ± 3.6

	Gesture Phase Seg. Proc.	first-order theorem prov.	mfeat-karhunen	mfeat-fourier	MiceProtein	satimage	steel-plates fault	Average Pos.-F1 (%)	Average Rank	Wins
DRIP-MPU (Ours)	29.8 ± 2.7	**22.5** ± 3.3	**92.4** ± 1.9	**77.2** ± 1.9	**99.1** ± 0.9	**75.0** ± 1.3	**68.3** ± 4.1	**61.9** ± 0.5	**1.960**	**17**
BBE-PU	0.0 ± 0.0	0.9 ± 1.2	8.1 ± 2.2	4.3 ± 1.8	4.0 ± 1.7	1.9 ± 2.1	43.2 ± 3.7	20.6 ± 1.9	9.760	1
LaGAM-PU	**30.6** ± 1.3	21.9 ± 3.8	81.0 ± 2.0	59.3 ± 2.3	73.1 ± 3.1	63.6 ± 2.0	49.2 ± 1.7	47.8 ± 0.9	4.160	2
CGenPU	17.6 ± 4.3	15.2 ± 2.6	34.5 ± 10.1	14.5 ± 5.1	21.4 ± 7.0	27.0 ± 9.3	47.1 ± 3.6	42.8 ± 1.4	4.560	2
VPU	0.2 ± 0.1	2.2 ± 0.8	2.6 ± 1.1	1.1 ± 0.6	0.1 ± 0.4	7.9 ± 5.0	21.3 ± 4.1	9.6 ± 1.0	13.320	0
PAN	0.0 ± 0.0	2.3 ± 3.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	3.2 ± 6.8	5.3 ± 6.8	25.7 ± 2.7	9.740	1
HolisticPU	11.5 ± 2.9	14.5 ± 2.9	2.1 ± 0.8	2.0 ± 0.9	0.2 ± 0.4	12.3 ± 7.9	6.3 ± 2.3	19.9 ± 0.8	9.480	0
P3Mix-C	5.5 ± 1.4	13.8 ± 2.9	8.6 ± 2.4	7.8 ± 2.1	8.4 ± 3.6	12.1 ± 2.6	32.6 ± 3.9	14.6 ± 1.3	10.040	0
P3Mix-E	4.0 ± 1.3	14.2 ± 2.5	6.5 ± 1.7	5.9 ± 1.0	6.0 ± 2.1	11.9 ± 2.6	33.3 ± 3.3	13.2 ± 1.1	11.200	0
PUET (KM2)	13.4 ± 4.6	14.6 ± 3.6	42.2 ± 4.1	38.0 ± 7.6	75.8 ± 6.2	70.6 ± 5.8	53.3 ± 4.8	38.4 ± 3.0	5.960	0
ScalePU (KM2)	3.1 ± 4.6	7.5 ± 2.3	1.4 ± 1.0	0.5 ± 0.8	4.0 ± 3.7	13.8 ± 7.0	33.5 ± 4.2	13.1 ± 2.1	12.000	0
nnPU (KM2)	0.0 ± 0.0	0.0 ± 0.0	0.6 ± 1.8	1.0 ± 2.2	5.2 ± 5.2	0.0 ± 0.0	8.3 ± 7.1	13.4 ± 3.1	12.580	0
Dist-PU (KM2)	11.3 ± 6.1	17.0 ± 4.1	80.7 ± 3.7	57.3 ± 4.0	79.0 ± 5.3	59.4 ± 6.2	46.8 ± 9.1	40.4 ± 2.2	6.040	1
PULDA (KM2)	10.1 ± 6.4	19.5 ± 3.9	17.5 ± 2.7	14.5 ± 4.1	69.2 ± 19.5	55.6 ± 5.9	58.9 ± 2.1	34.1 ± 3.1	6.760	1
UREA (KM2)	10.4 ± 2.3	14.3 ± 3.9	17.6 ± 10.1	19.9 ± 5.7	13.5 ± 5.9	49.8 ± 7.9	22.9 ± 7.2	33.3 ± 2.3	6.760	0
AREA (KM2)	3.1 ± 1.5	10.5 ± 2.9	12.2 ± 1.7	9.2 ± 1.5	0.2 ± 0.4	32.5 ± 3.9	33.6 ± 6.3	12.1 ± 0.8	11.680	0
PUET (real)	7.7 ± 1.7	15.7 ± 2.8	31.4 ± 3.9	26.9 ± 4.8	53.5 ± 4.7	73.3 ± 2.2	60.7 ± 4.1	50.9 ± 0.9	–	–
ScalePU (real)	0.5 ± 0.4	7.9 ± 2.5	1.3 ± 0.6	0.5 ± 0.7	3.6 ± 3.6	11.7 ± 4.6	33.8 ± 5.2	14.8 ± 1.7	–	–
nnPU (real)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	10.1 ± 7.1	24.8 ± 1.7	–	–
Dist-PU (real)	3.4 ± 1.4	18.5 ± 3.6	78.3 ± 2.2	53.2 ± 3.2	75.1 ± 4.0	63.8 ± 3.2	53.6 ± 3.1	52.2 ± 1.6	–	–
PULDA (real)	2.1 ± 1.1	20.8 ± 3.3	11.9 ± 3.5	10.3 ± 2.3	61.4 ± 7.9	57.6 ± 4.4	63.1 ± 3.1	44.1 ± 1.7	–	–
UREA (real)	11.4 ± 0.7	12.2 ± 4.0	28.1 ± 7.2	23.3 ± 5.4	13.7 ± 5.3	55.1 ± 5.5	24.5 ± 6.4	42.4 ± 1.5	–	–
AREA (real)	3.0 ± 1.3	11.1 ± 2.1	12.0 ± 1.9	8.2 ± 1.4	0.3 ± 0.5	32.5 ± 5.2	34.4 ± 3.4	13.2 ± 0.8	–	–

Table 11: AUC (%) results on all 25 datasets in the 15-shot setting.

Method	Australian	blood-transfusion	climate-model sim.-crashes	credit-approval	credit-g	cylinder-bands	diabetes	dresses-sales	kc2
DRIP-MPU (Ours)	**85.9** ± 5.1	66.1 ± 7.2	53.1 ± 9.4	**89.4** ± 1.4	**69.9** ± 6.0	66.7 ± 8.5	**73.0** ± 4.6	51.0 ± 7.7	66.2 ± 11.5
BBE-PU	77.8 ± 6.2	**71.7** ± 4.8	**60.9** ± 12.1	80.7 ± 4.9	59.7 ± 5.2	57.6 ± 7.7	72.3 ± 4.6	54.9 ± 5.6	70.9 ± 3.8
LaGAM-PU	67.6 ± 13.3	66.4 ± 3.5	53.8 ± 13.1	79.9 ± 7.0	58.2 ± 5.9	56.5 ± 7.9	68.5 ± 8.8	55.3 ± 6.6	63.4 ± 8.5
CGenPU	72.1 ± 14.6	61.8 ± 5.0	57.1 ± 13.4	70.8 ± 6.7	61.8 ± 5.7	65.6 ± 5.5	64.6 ± 9.8	53.2 ± 4.9	57.5 ± 8.1
VPU	62.8 ± 7.4	63.2 ± 8.8	57.2 ± 13.7	70.9 ± 6.2	53.6 ± 4.4	55.8 ± 4.4	60.8 ± 6.6	54.3 ± 7.7	60.4 ± 10.5
PAN	85.0 ± 4.5	67.3 ± 9.4	49.7 ± 17.4	81.1 ± 6.8	53.4 ± 4.4	63.1 ± 14.0	72.6 ± 5.4	55.0 ± 4.8	**74.9** ± 0.8
HolisticPU	24.3 ± 7.8	55.7 ± 10.8	44.7 ± 14.9	24.2 ± 6.0	40.8 ± 4.0	45.7 ± 10.2	33.5 ± 8.4	45.0 ± 6.0	40.1 ± 14.0
P3Mix-C	60.8 ± 5.6	59.7 ± 4.3	56.9 ± 9.1	71.4 ± 6.2	56.7 ± 4.8	59.0 ± 6.7	66.4 ± 5.6	53.7 ± 7.6	64.0 ± 9.7
P3Mix-E	68.0 ± 8.1	62.0 ± 4.4	58.4 ± 11.9	74.8 ± 5.8	56.7 ± 4.2	58.1 ± 7.5	67.2 ± 5.2	53.9 ± 5.9	58.9 ± 8.8
PUET (KM2)	79.0 ± 8.8	57.5 ± 7.0	53.7 ± 10.4	78.0 ± 7.7	60.8 ± 9.2	**69.5** ± 9.9	70.7 ± 4.1	51.8 ± 6.1	71.3 ± 3.7
ScalePU (KM2)	73.9 ± 9.1	63.6 ± 7.4	54.1 ± 10.2	76.9 ± 6.1	60.3 ± 5.5	62.4 ± 6.0	69.2 ± 9.4	54.8 ± 6.2	67.5 ± 6.9
nnPU (KM2)	68.7 ± 11.8	55.9 ± 5.4	55.0 ± 13.2	69.3 ± 10.1	50.3 ± 7.8	55.6 ± 14.5	58.4 ± 12.9	**56.2** ± 4.8	72.0 ± 7.5
Dist-PU (KM2)	79.5 ± 13.9	67.8 ± 7.4	45.3 ± 6.7	81.0 ± 9.6	57.0 ± 9.9	61.9 ± 12.0	71.3 ± 5.0	51.8 ± 7.3	74.1 ± 0.8
PULDA (KM2)	77.8 ± 8.4	69.6 ± 4.0	50.9 ± 5.9	78.7 ± 8.5	60.0 ± 3.3	62.3 ± 9.5	69.5 ± 5.0	51.4 ± 7.1	68.4 ± 4.2
UREA (KM2)	60.9 ± 4.9	59.6 ± 10.1	56.3 ± 12.5	63.7 ± 5.6	57.7 ± 4.0	59.4 ± 2.3	60.9 ± 5.2	53.6 ± 6.0	59.6 ± 9.5
AREA (KM2)	68.7 ± 11.4	59.0 ± 8.5	53.3 ± 14.8	69.8 ± 7.1	58.4 ± 5.2	62.4 ± 4.5	58.6 ± 6.4	53.7 ± 7.1	51.0 ± 11.2
PUET (real)	84.0 ± 5.1	59.3 ± 8.1	53.5 ± 11.3	84.7 ± 3.8	65.5 ± 4.2	72.3 ± 8.8	73.2 ± 4.2	53.3 ± 7.3	71.5 ± 4.0
ScalePU (real)	73.7 ± 9.5	63.5 ± 7.5	54.2 ± 10.3	77.1 ± 5.9	60.3 ± 5.6	62.2 ± 6.0	69.1 ± 9.4	54.7 ± 6.3	67.0 ± 7.4
nnPU (real)	83.3 ± 6.2	61.2 ± 4.9	54.8 ± 14.3	82.4 ± 4.9	52.8 ± 5.2	65.3 ± 10.7	73.9 ± 4.2	55.2 ± 5.8	74.0 ± 1.4
Dist-PU (real)	81.9 ± 7.9	72.7 ± 4.5	46.3 ± 6.6	80.0 ± 6.1	61.5 ± 5.0	64.8 ± 9.9	73.4 ± 4.9	54.6 ± 5.5	74.0 ± 0.9
PULDA (real)	82.0 ± 7.8	69.9 ± 3.6	47.6 ± 8.9	80.9 ± 6.6	58.9 ± 4.0	67.4 ± 4.9	71.0 ± 5.9	51.2 ± 5.5	68.4 ± 4.1
UREA (real)	69.6 ± 6.2	62.2 ± 8.4	55.3 ± 11.1	71.2 ± 4.9	57.3 ± 4.3	60.4 ± 4.0	63.7 ± 4.0	50.5 ± 4.1	59.1 ± 7.1
AREA (real)	69.0 ± 12.7	61.4 ± 5.9	53.5 ± 14.9	71.6 ± 8.0	57.3 ± 6.0	61.6 ± 4.2	62.2 ± 8.1	51.2 ± 6.9	53.1 ± 9.7

	pc3	segment	sonar	spambase	tic-tac-toe	wdbc	analcatdata_author	analcatdata_dmft	car
DRIP-MPU (Ours)	72.9 ± 8.4	99.0 ± 1.2	78.9 ± 10.1	**89.8** ± 3.8	55.9 ± 4.6	98.7 ± 1.4	**100.0** ± 0.0	51.6 ± 1.5	90.4 ± 1.9
BBE-PU	75.9 ± 5.0	96.5 ± 1.9	75.5 ± 9.8	61.3 ± 8.3	55.2 ± 4.8	96.3 ± 1.5	94.4 ± 1.9	51.8 ± 1.6	80.2 ± 7.8
LaGAM-PU	69.5 ± 7.4	97.0 ± 1.3	71.5 ± 9.3	54.6 ± 5.6	55.7 ± 7.8	93.7 ± 5.2	97.3 ± 0.6	**53.3** ± 2.8	85.4 ± 2.3
CGenPU	56.9 ± 10.4	75.2 ± 11.3	64.9 ± 11.8	57.2 ± 9.7	55.1 ± 7.2	89.4 ± 7.7	80.9 ± 6.4	52.8 ± 3.5	78.9 ± 3.4
VPU	64.8 ± 4.1	96.5 ± 2.2	71.7 ± 7.8	65.7 ± 2.2	53.4 ± 3.3	82.4 ± 4.0	77.7 ± 1.7	51.6 ± 2.5	91.6 ± 2.6
PAN	68.2 ± 19.8	98.0 ± 3.1	53.3 ± 9.9	68.4 ± 3.8	54.6 ± 4.8	**99.5** ± 0.3	87.8 ± 5.0	51.7 ± 3.0	53.3 ± 4.2
HolisticPU	30.8 ± 9.7	35.4 ± 9.3	30.3 ± 5.8	67.3 ± 2.8	47.0 ± 6.2	5.3 ± 3.4	8.4 ± 3.9	50.7 ± 1.9	26.7 ± 5.0
P3Mix-C	68.1 ± 4.8	96.9 ± 1.9	65.2 ± 10.4	68.9 ± 4.6	55.2 ± 4.6	81.7 ± 3.4	76.5 ± 2.8	51.8 ± 1.7	91.7 ± 1.7
P3Mix-E	64.2 ± 6.7	99.1 ± 0.5	67.0 ± 9.9	63.9 ± 4.7	55.8 ± 4.4	86.5 ± 5.2	72.5 ± 3.2	51.3 ± 1.8	89.5 ± 1.4
PUET (KM2)	74.5 ± 3.9	98.3 ± 2.0	73.1 ± 12.4	79.7 ± 6.5	**57.3** ± 4.2	97.6 ± 2.2	99.1 ± 0.5	50.8 ± 2.2	87.2 ± 5.4
ScalePU (KM2)	67.3 ± 3.9	97.5 ± 1.3	**82.8** ± 6.5	67.6 ± 4.4	56.2 ± 4.2	97.3 ± 2.4	89.5 ± 1.4	50.9 ± 2.4	85.6 ± 4.6
nnPU (KM2)	70.7 ± 6.5	76.1 ± 5.1	72.1 ± 5.1	50.4 ± 2.3	53.1 ± 6.1	98.6 ± 1.0	91.9 ± 2.4	50.3 ± 2.3	66.0 ± 8.4
Dist-PU (KM2)	**77.6** ± 3.0	97.6 ± 1.7	73.5 ± 11.0	61.9 ± 4.4	55.4 ± 4.7	98.5 ± 1.3	99.0 ± 0.7	51.2 ± 2.0	86.5 ± 11.2
PULDA (KM2)	72.2 ± 7.7	**99.3** ± 0.5	72.5 ± 13.2	57.3 ± 4.2	54.9 ± 6.1	94.9 ± 2.9	92.9 ± 3.2	52.1 ± 2.4	92.4 ± 2.1
UREA (KM2)	57.1 ± 3.3	93.3 ± 4.6	65.6 ± 8.1	60.2 ± 3.9	52.8 ± 4.1	79.2 ± 7.3	63.0 ± 4.7	51.4 ± 2.7	73.5 ± 5.6
AREA (KM2)	59.1 ± 7.9	95.6 ± 2.3	70.7 ± 9.5	59.3 ± 7.3	53.0 ± 7.1	87.8 ± 3.9	75.1 ± 4.2	51.0 ± 1.9	**93.4** ± 1.6
PUET (real)	71.4 ± 4.7	99.7 ± 0.4	79.2 ± 7.2	88.8 ± 1.6	58.7 ± 5.1	98.8 ± 0.9	99.3 ± 0.2	50.9 ± 1.7	83.9 ± 3.2
ScalePU (real)	67.7 ± 3.9	97.6 ± 1.2	81.9 ± 8.0	67.3 ± 4.3	56.3 ± 4.0	97.1 ± 2.3	89.3 ± 2.2	50.9 ± 2.5	84.7 ± 4.9
nnPU (real)	66.6 ± 3.3	84.4 ± 3.2	79.3 ± 6.7	58.5 ± 2.4	55.9 ± 3.7	99.3 ± 0.4	90.1 ± 1.7	50.3 ± 1.9	53.7 ± 2.1
Dist-PU (real)	77.3 ± 3.3	98.4 ± 1.3	81.0 ± 6.0	64.8 ± 7.5	56.4 ± 4.7	99.2 ± 0.5	99.5 ± 0.3	51.6 ± 2.5	72.9 ± 7.3
PULDA (real)	71.1 ± 8.2	99.5 ± 0.3	49.5 ± 1.4	50.5 ± 4.0	53.0 ± 4.1	96.7 ± 1.9	95.2 ± 1.1	52.4 ± 2.4	93.4 ± 2.8
UREA (real)	56.0 ± 2.8	91.7 ± 4.2	73.3 ± 8.4	69.9 ± 2.1	54.3 ± 4.6	84.8 ± 5.3	61.8 ± 2.4	51.8 ± 2.2	66.2 ± 1.7
AREA (real)	57.0 ± 4.8	96.5 ± 2.1	71.7 ± 9.8	63.8 ± 5.2	52.9 ± 7.2	88.3 ± 3.9	75.6 ± 4.5	51.1 ± 1.6	93.1 ± 2.1

	Gesture Phase Seg. Proc.	first-order theorem prov.	mfeat-karhunen	mfeat-fourier	MiceProtein	satimage	steel-plates fault	Average AUC (%)	Average Rank	Wins
DRIP-MPU (Ours)	58.8 ± 2.4	59.9 ± 2.0	**99.3** ± 0.4	**97.2** ± 0.6	**99.5** ± 0.8	**96.4** ± 0.3	89.5 ± 1.5	**78.4** ± 1.2	**3.560**	**10**
BBE-PU	56.1 ± 4.0	55.0 ± 3.0	91.9 ± 1.2	83.9 ± 1.7	94.0 ± 1.1	74.4 ± 4.4	84.2 ± 2.4	73.3 ± 1.7	6.520	2
LaGAM-PU	58.3 ± 1.5	**61.5** ± 2.5	95.8 ± 0.7	91.4 ± 0.7	93.4 ± 0.9	92.8 ± 0.5	85.4 ± 0.9	73.0 ± 1.2	6.840	2
CGenPU	54.6 ± 2.4	55.8 ± 2.7	78.9 ± 4.8	66.0 ± 2.6	70.1 ± 3.9	65.9 ± 7.5	81.5 ± 1.5	66.0 ± 1.2	10.600	0
VPU	58.0 ± 1.8	60.3 ± 1.6	90.0 ± 1.0	83.8 ± 0.9	90.9 ± 1.0	88.0 ± 1.1	87.3 ± 1.0	70.1 ± 1.0	8.800	0
PAN	57.7 ± 6.2	56.3 ± 2.5	90.0 ± 1.2	73.7 ± 1.9	89.8 ± 1.3	84.2 ± 3.1	84.0 ± 2.9	70.9 ± 1.7	7.400	2
HolisticPU	48.3 ± 3.0	51.4 ± 3.0	11.7 ± 1.7	14.9 ± 1.4	11.0 ± 1.7	38.8 ± 11.7	28.6 ± 3.6	34.4 ± 1.0	15.520	0
P3Mix-C	57.5 ± 2.8	58.4 ± 1.9	81.4 ± 1.8	77.6 ± 2.4	80.1 ± 2.9	85.4 ± 1.4	84.7 ± 1.3	69.2 ± 1.2	9.040	0
P3Mix-E	56.6 ± 2.6	58.1 ± 2.2	77.7 ± 1.2	73.6 ± 2.2	85.0 ± 3.1	83.9 ± 1.8	84.6 ± 1.7	69.1 ± 1.5	9.280	0
PUET (KM2)	**61.6** ± 4.6	60.4 ± 2.3	96.1 ± 1.0	93.1 ± 0.6	97.7 ± 0.9	95.0 ± 0.9	88.6 ± 1.1	76.1 ± 1.5	4.960	3
ScalePU (KM2)	50.9 ± 2.1	56.4 ± 1.4	88.0 ± 1.5	82.6 ± 1.8	93.4 ± 1.2	83.8 ± 1.7	85.8 ± 1.5	72.7 ± 1.1	7.360	1
nnPU (KM2)	54.5 ± 3.7	54.0 ± 1.5	80.5 ± 13.2	79.2 ± 9.0	69.4 ± 3.0	69.3 ± 2.4	75.8 ± 6.6	66.1 ± 1.9	11.120	1
Dist-PU (KM2)	58.5 ± 2.4	60.5 ± 2.0	97.8 ± 0.7	92.9 ± 1.0	97.5 ± 0.8	94.4 ± 0.8	89.0 ± 1.1	75.3 ± 1.5	5.320	1
PULDA (KM2)	55.7 ± 4.2	60.7 ± 2.5	96.3 ± 1.4	89.5 ± 1.4	98.2 ± 1.4	91.0 ± 1.0	**90.2** ± 0.6	74.3 ± 0.9	5.880	2
UREA (KM2)	52.4 ± 3.7	56.7 ± 2.8	58.6 ± 5.0	61.9 ± 4.9	56.0 ± 2.2	77.7 ± 6.2	59.4 ± 4.0	62.0 ± 1.0	12.400	0
AREA (KM2)	44.1 ± 1.3	56.0 ± 1.8	66.8 ± 3.2	66.7 ± 2.6	63.7 ± 1.6	86.4 ± 2.1	77.9 ± 2.7	65.7 ± 1.4	11.400	1
PUET (real)	66.9 ± 1.7	60.6 ± 2.3	96.2 ± 0.9	92.3 ± 0.9	98.5 ± 0.7	95.4 ± 0.6	89.4 ± 0.7	77.9 ± 0.8	–	–
ScalePU (real)	50.7 ± 1.8	56.2 ± 1.7	87.9 ± 1.6	82.7 ± 1.8	93.5 ± 1.3	83.0 ± 2.0	85.8 ± 1.1	72.6 ± 1.2	–	–
nnPU (real)	54.0 ± 1.7	53.3 ± 1.4	92.7 ± 0.4	86.9 ± 2.0	88.2 ± 0.8	69.2 ± 2.7	72.5 ± 3.3	70.3 ± 1.2	–	–
Dist-PU (real)	52.2 ± 2.1	61.0 ± 2.3	98.2 ± 0.4	92.7 ± 1.3	97.8 ± 0.7	94.7 ± 0.6	89.2 ± 1.3	75.8 ± 1.0	–	–
PULDA (real)	50.9 ± 2.8	61.1 ± 2.2	96.1 ± 1.0	88.6 ± 1.3	99.1 ± 0.4	90.8 ± 1.0	90.2 ± 0.6	73.4 ± 0.6	–	–
UREA (real)	52.4 ± 2.4	54.5 ± 2.7	64.3 ± 5.0	64.8 ± 4.5	58.1 ± 2.4	79.6 ± 4.1	60.7 ± 3.0	63.8 ± 1.5	–	–
AREA (real)	43.6 ± 0.9	56.4 ± 1.5	65.8 ± 2.3	66.4 ± 2.5	63.3 ± 2.6	85.2 ± 2.3	77.5 ± 3.1	66.0 ± 1.0	–	–