Supplementary: Method

1 Federated binning

Algorithm 1 Federated Binning Workflow

```
Input: Local Data $D_i$ across $C$ clients, Number of Bins $B$
for each client $i \in \{1, 2, \ldots, C\}$ do
    Compute local non-zero values in $D_i$
    Compute local bin edges $B_i \leftarrow \text{Quantile}(D_i, B)$
    Send $B_i$ and local sample size $n_i$ to the coordinator
end for
$B\_global \leftarrow \frac{\sum_{i=1}^{C} n_i \times B_i}{\sum_{i=1}^{C} n_i}$    ▷ Weighted average of bin edges
for each client $i \in \{1, 2, \ldots, C\}$ do
    Apply $B\_global$ to $D_i$
    $D_i^{binned} \leftarrow \text{Digitize}(D_i, B\_global)$
end for
Output: Global bin edges $B\_global$, Binned data $D_i^{binned}$ for each client
```

2 Federated reference mapping

Algorithm 2 Federated Zero-Shot Reference Mapping

```
Input: $\{\mathcal{R}_i\}_{i=1}^{C}$, $\mathcal{Q}$, $\mathcal{F}$, $k$
$\mathcal{Q}' \leftarrow \mathcal{F}(\mathcal{Q})$
for each client $i \in \{1, \ldots, C\}$ do
    $\mathcal{R}'_i \leftarrow \mathcal{F}(\mathcal{R}_i)$
    $(\mathcal{D}_i, \mathcal{I}_i) \leftarrow \|\mathcal{Q}' - \mathcal{R}'_i\|$    ▷ keep top-$k$ per client
end for
$k\_nearest_j \leftarrow \text{TopK}\big(\{\mathcal{I}_i[j, 1 \ldots k]\}_{i=1}^{C}\big) \quad (\forall j \in \mathcal{Q})$
for each client $i \in \{1, \ldots, C\}$ do
    $\mathcal{V}_i[j] \leftarrow \mathcal{L}_i\big[\mathcal{I}_i[j]\big] \quad (\forall j \in k\_nearest_j)$
end for
$Predictions \leftarrow \text{mode}\{\mathcal{V}_i[j]\}_{i=1}^{C}$ for each $j \in \mathcal{Q}$
```

2.1 Secure federated reference mapping

Algorithm 3 Secure Federated Zero-Shot Reference Mapping

Definitions:
  $\mathcal{P} \leftarrow$ number of SMPC parties
  $\langle \cdot \rangle \leftarrow [\![\cdot]\!]_{\mathcal{P}}$ (secret-sharing)
  $(\!|\cdot|\!)_{\mathcal{P}}$ (SMPC reveal)
  $\Lambda \leftarrow$ large masking constant
  $\Phi\colon \mathcal{L} \to \{1, \ldots, |\mathcal{L}|\}$ (label$\to$index map)
  shapes: $\langle \mathcal{D} \rangle \in \mathbb{F}^{|\mathcal{Q}| \times (Ck)}, \ \ldots$
Input: query Q, references $\{\mathcal{R}_i\}_{i=1}^{C}$, model $\mathcal{F}$, neighbor count k
$\langle \mathcal{Q} \rangle \leftarrow [\![\mathcal{F}(\mathcal{Q})]\!]_{\mathcal{P}}$ ▷ Secret-share embedded query
$\mathcal{O} \leftarrow (\!|\sum_{i=1}^{C} \langle n_i \rangle|\!)_{\mathcal{P}}$ ▷ Reveal total number of samples as offset
$\mathcal{L} \ = \ \bigcup_{i=1}^{C} l_i$ ▷ All unique labels
$\phi = \left[\Phi(\ell) \ : \ \ell \in \mathcal{L}\right]$ ▷ label index-vector
$\langle \mathcal{D} \rangle \leftarrow \mathbf{0}^{\,|\mathcal{Q}| \times (C\,k)}, \quad \langle \mathcal{I} \rangle \leftarrow \mathbf{0}^{\,|\mathcal{Q}| \times (C\,k)}$
for $i = 1$ **to** C **do** ▷ Local secure k-NN per client
  ClientSetup($\mathcal{F}$, Φ, $(i-1)\mathcal{O}$)
  $(\langle d \rangle, \langle ind \rangle) \leftarrow$ SecureKNN($\langle \mathcal{Q} \rangle, k$)
  $\langle \mathcal{D} \rangle_{:,(i-1)*k:i*k} \leftarrow \langle d \rangle, \quad \langle \mathcal{I} \rangle_{:,(i-1)*k:i*k} \leftarrow \langle ind \rangle$
end for
$\langle \mathcal{N} \rangle \leftarrow \mathbf{0}^{\,|\mathcal{Q}| \times k}$
for $j = 1$ **to** k **do**
  $\langle \mathcal{M} \rangle = \mathrm{minMask}(\langle \mathcal{D} \rangle)$
  $\langle \mathcal{N} \rangle_{:,j} = \langle \mathcal{M} \rangle \, \langle \mathcal{I} \rangle$
  $\langle \mathcal{D} \rangle \mathrel{+}= \ \Lambda \, \langle \mathcal{M} \rangle$
end for
for $i = 1$ **to** C **do**
  $\langle V_i \rangle = \mathrm{SecureVote}(\langle \mathcal{N} \rangle, |\mathcal{Q}|, |\mathcal{L}|)$
end for
$\langle \mathcal{V} \rangle \leftarrow \sum_{i=1}^{C} \langle V_i \rangle$
$\langle M \rangle \leftarrow \mathrm{maxMask}(\langle \mathcal{V} \rangle)$ ▷ secret-shared one-hot of highest-vote class
$M^{\text{plain}} \leftarrow (\!|\langle M \rangle|\!)_{\mathcal{P}}$
predictions $= \sum_{\ell=1}^{|\mathcal{L}|} M_{:,\ell}^{\text{plain}} \, \phi_\ell$
return predictions

Algorithm 4 ClientSetup

Precondition: $\mathcal{R}$, l, $\mathcal{P}$
Input: Model $\mathcal{F}$, Offset $\mathcal{O}$, label-to-index map Φ
$\langle \mathcal{R} \rangle \leftarrow [\![\mathcal{F}(\mathcal{R})]\!]_{\mathcal{P}}$ ▷ Secret-share reference embeddings
$\langle r \rangle \leftarrow [\![[0, 1, \ldots, |\mathcal{R}| - 1] + \mathcal{O}]\!]_{\mathcal{P}}$ ▷ Secret-share index vector
$\ell \leftarrow [\Phi(l_r) : \forall r \in \{0, \ldots, |R| - 1\}]$ ▷ map local labels
$\langle \ell \rangle \leftarrow [\![\ell]\!]_{\mathcal{P}}$ ▷ secret-share mapped label indices
$\Lambda \leftarrow$ large mask constant

Algorithm 5 SecureKNN

Precondition: $\langle \mathcal{R} \rangle$, $\langle r \rangle$, Λ
Input: secret-shared query $\langle \mathcal{Q} \rangle$, neighbor count k
$\langle \mathcal{D}^{(0)} \rangle = (\langle \mathcal{Q} \rangle \circ \langle \mathcal{Q} \rangle)\, \mathbf{1}_{|\mathcal{R}|}^{T} + \mathbf{1}_{|\mathcal{Q}|}\, (\langle \mathcal{R} \rangle \circ \langle \mathcal{R} \rangle)^{T} - 2\, \langle \mathcal{Q} \rangle\, \langle \mathcal{R} \rangle^{T}$
$\langle \mathcal{T} \rangle \leftarrow \mathbf{0}^{\,|\mathcal{Q}| \times k}, \quad \langle \mathcal{I} \rangle \leftarrow \mathbf{0}^{\,|\mathcal{Q}| \times k}$
for $j = 1$ **to** k **do**
    $\langle \mathcal{M} \rangle = \mathrm{minMask}(\langle \mathcal{D}^{(j-1)} \rangle)$ ▷ One-hot argmin
    $\langle \mathcal{T} \rangle_{:,j} = (\langle \mathcal{M} \rangle \odot \langle \mathcal{D}^{(0)} \rangle)\, \mathbf{1}_{|\mathcal{R}|}$ ▷ Hadamard-product
    $\langle \mathcal{I} \rangle_{:,j} = \langle \mathcal{M} \rangle\, \langle r \rangle$
    $\langle \mathcal{D}^{(j)} \rangle = \langle \mathcal{D}^{(j-1)} \rangle + \Lambda\, \langle \mathcal{M} \rangle$ ▷ Suppress the chosen minima
end for
return $(\langle \mathcal{T} \rangle,\ \langle \mathcal{I} \rangle)$

Algorithm 6 SecureVote

Precondition: $\langle r \rangle$, $\langle \ell \rangle$
Input: Global neighbors $\langle \mathcal{N} \rangle$, number of query samples n_q,number of unique labels n_l.
$\langle V \rangle \leftarrow \mathbf{0}^{\,n_q \times n_l}$
for $j = 1$ **to** k **do**
    $\langle M \rangle \leftarrow \mathrm{eqMask}(\,\langle \mathcal{N} \rangle_{:,j}, \langle r \rangle)$ ▷ secret-shared equality check
    $\langle v \rangle \leftarrow \langle M \rangle\, \langle \ell \rangle$
    $\langle V \rangle \mathrel{+}= \ \mathrm{oneHot}(\langle v \rangle,\ n_l)$ ▷ $\langle v \rangle = 0 \rightarrow$ *no vote*
end for
return $\langle V \rangle$

Notation	Description
k	Number of nearest neighbors.
C	Number of clients (parties) in the federation.
i	Client i index.
$\mathcal{F}$	Shared embedding model.
$[\![\cdot]\!]_{\mathcal{P}}$	Secret-sharing operator among $\mathcal{P}$ parties.
$\mathcal{P}$	Number of SMPC parties.
$\langle\cdot\rangle$	Secret-sharing operand under the SMPC protocol.
$\mathcal{Q}$	Query dataset.
$\mathcal{R}$	Reference dataset.
$\mathcal{Q}'$	Embedded query.
$\langle\mathcal{Q}\rangle$	Secret-shared embedded query.
$\mathcal{R}'$	Embedded reference dataset.
$\langle\mathcal{R}\rangle$	Secret-shared reference data.
l_i	Client-i local labels.
$\mathcal{L}$	Global set of all unique labels.
$\mathcal{O}$	Global index offset.
Φ	Bijection from label strings to integer indices.
ϕ	Label index-vector.
r_i	Client-i offset indices.
$\langle r_i\rangle$	Secret-shared offset indices of client-i.
ℓ_i	Client-i mapped local labels.
$\langle\ell_i\rangle$	Secret-shared mapped label indices of client-i.
Λ	Large constant value.
$\langle\mathcal{D}^{(0)}\rangle$	Initial local distance matrix before masking.
$\langle\mathcal{D}^{(j)}\rangle$	local distance matrix after masking out the first j minima.
$\langle\mathcal{T}\rangle_{:,j}$	Distance vector for j-st nearest neighbor per query.
$\langle\mathcal{I}\rangle_{:,j}$	Index vector for j-st nearest neighbor per query.
$\langle\mathcal{D}\rangle$	Secret-shared query-reference distance matrix.
$\langle\mathcal{I}\rangle$	Secret-shared indices of reference samples in $\langle\mathcal{D}\rangle$.
$\langle\mathcal{N}\rangle$	One-hot secret indices of the global nearest neighbor.
$\langle v\rangle$	Client's secret-shared vote vector (per neighbor).
$\langle V_i\rangle$	Client-i's secret-shared vote matrix.
$\langle\mathcal{V}\rangle$	Aggregated secret vote matrix across all clients.
M^{plain}	One-hot matrix of indices of the most frequent votes.

Table 1: Notation table.