

When Masking Meets Sparsity: COVERT for Private, Communication-Efficient Vertical Federated Tensor Regression

Axel Faes^{a,*}, Maureen M. J. Guichelaar^b, Donald E. Bouman^c, Maryam Amir Haeri^a

^aDepartment of Learning, Data Analytics and Technology, University of Twente, Enschede, The Netherlands

^bDepartment of Gastroenterology and Hepatology, Medisch Spectrum Twente, Enschede, The Netherlands

^cDepartment of Radiology, Medisch Spectrum Twente, Enschede, The Netherlands

Abstract

Cross-institutional biomedical analysis is often *vertically* partitioned: the same patients are described by modalities held by different parties (an electrocardiogram at one system, laboratory panels at another, imaging at a third), and one party holds the outcome label. Existing vertical federated learning (VFL) methods are linear: privacy-preserving partial least squares (PLS) shares a sample-mode score and keeps each party’s feature loadings private. Biomedical signals are multiway, and block-term tensor regression (BTTR) exploits that structure with sparse, interpretable components. We introduce *COVERT* (coupled vertical tensor regression), the supervised, multilinear, vertical-federated generalization of PLS: parties fit one coupled rank- $(L_r, L_r, 1)$ block-term regression with a shared sample-mode score and strictly private feature-mode factors. *COVERT matches* the centralized model (losslessly) and linear vertical PLS in accuracy, and adds (i) a differential-privacy guarantee with noise *independent of the feature dimension*; (ii) *dimension-free communication*: only the cohort-sized score and residual cross parties, so per-round traffic stays flat as local representations grow; and (iii) *sparse, interpretable multiway components* stable across resampling. A no-go characterization shows the design is forced: among orthogonal privacy masks only signed permutations preserve a sparse decomposition’s support, so the masked-decomposition trick behind linear VFL cannot recover a sparse core, and the forced sparsity-preserving route provably matches centralized fitting. On a 3-party MIMIC-IV-ECG cohort *COVERT* reaches 0.745 AUROC, statistically equivalent to centralized fitting and linear vertical PLS, cuts per-round communication by an order of magnitude, keeps an 87%-sparse interpretable core, and adds differential privacy that degrades gracefully.

Keywords: Vertical federated learning, differential privacy, tensor decomposition, block-term regression, interpretable machine learning, communication efficiency, electrocardiography

1. Introduction

Much of the highest-value data in medicine is *vertically* partitioned. A single patient’s record is assembled from organizationally separate sources: an electrocardiogram (ECG) captured by a cardiology system, laboratory panels stored in a hospital information system, imaging held by radiology, each governed by its own access controls and, increasingly, by its own institution. The entities (patients) overlap; the *features* do not. Pooling the views centrally would maximize predictive power, but is precisely what privacy regulation and institutional policy forbid.

Vertical federated learning (VFL) targets exactly this regime (shared samples, disjoint features, labels held by a single “active” party) [1]. Yet the federated learning literature is overwhelmingly *horizontal* (parties hold different patients, the same features) [2, 3, 4]; vertical methods are comparatively few, and the principled ones are *linear and matrix-based*. The closest precedent, privacy-preserving partial least squares (P3LS) [5], fits a PLS regression across parties by sharing a sample-mode latent score while each party keeps its feature loadings private, using a masked singular value decomposition (SVD) for privacy.

*Corresponding author.

Email address: axel.faes@utwente.nl (Axel Faes)

Biomedical signals, however, are intrinsically *multiway*. A multi-lead ECG is a channels $\times$ time ($\times$ beats) tensor; longitudinal imaging is voxels $\times$ time; electronic-health-record panels are features $\times$ time. Multilinear models that respect this structure, in particular block-term tensor regression (BTTR) [6, 7], fit a small number of low-multilinear-rank components that are both predictive and interpretable, and do so with automatic, *sparse* component selection. No vertical federated method exploits this structure: the only block-term work in a federated setting is our own *horizontal* FBTTT [4], which averages factor matrices across parties holding the same features and is structurally inapplicable to the vertical case.

This paper introduces **COVERT**, *coupled vertical tensor regression*: the supervised, multilinear, vertical generalization of federated PLS. Parties jointly fit one coupled rank- $(L_r, L_r, 1)$ block-term regression in which the sample-mode score of each component is shared while every party’s feature-mode factors remain private; the active party holds the labels. COVERT builds on the block-term tensor regression line [6, 7] and its horizontal federation [4], but neither applies here: both share or average feature-mode factors across parties, which is exactly what the vertical setting forbids.

The difficulty is not the model but its privacy. P3LS is private *because* PLS reduces to an SVD of a cross-product matrix, and orthogonal masks pass through an SVD unchanged. BTTR does not reduce to a plain SVD: its component-selection routine (Automatic Component Extraction, ACE) fits a *sparse* Tucker decomposition of a cross-covariance tensor, using soft-thresholding and rank pruning. Orthogonal masking does not commute with sparsity (soft-thresholding in a rotated basis is not the rotation of soft-thresholding), so the masked-SVD recipe cannot reproduce ACE’s sparse core. *The privacy mask fights the sparse-core pruning*. We turn this tension into a theory of the design space, contributing:

1. **COVERT, a multilinear vertical-federated model** (§4): the supervised, sparse, block-term generalization of federated PLS, fit by securely aggregating only the shared sample-mode score while every party’s feature-mode factors stay local. *It matches* centralized coupled BTTR and the linear vertical-PLS baseline in accuracy (§5). Accuracy is at parity by design; the contribution is what the design buys around it.
2. **Dimension-free privacy and communication** (§4.8): because only the cohort-sized score and residual cross the wire, both the differential-privacy noise and the per-round communication are *independent of the feature-mode dimension*: privacy and bandwidth stay flat no matter how rich each party’s local representation is, where the masked-dense alternative pays in the ambient dimension.
3. **Interpretable multiway components** (§5): the sparse rank- $(L_r, L_r, 1)$ blocks are per-modality, per-party signatures; on a genuinely multiway party they are reproducible far above a label-permutation null, structure the linear, score-only baselines do not expose.
4. **A design-space theory** (§4.4; *full proofs in Appendix Appendix A*): a *no-go* characterization (only signed-permutation masks preserve a sparse decomposition’s support, so masked-decomposition privacy is incompatible with sparse-core recovery, *forcing* the local-decompose-share-the-score route, Cor. 1) and an *exactness* result (that route is exact block-coordinate descent on the coupled objective (4), so federation costs no accuracy).

We are explicit about what we do not claim. “Each modality is a party” is an established VFL setup [1, 8]; our contribution is the *model class*, not the setup. Aligning or sharing a sample-mode score across parties is, in the linear case, exactly what P3LS does; our contribution is its multilinear, sparse, block-term form and the resolution of the obstruction that the linear case never faces.

2. Related Work

Federated tensor factorization (horizontal). Federated CP/PARAFAC factorization is mature, from TRIP [3] through generalized [9] and PARAFAC2 [10] variants; coupled tensor-train federation shares feature-mode factors while privatizing the *sample* mode [11]. All split the sample mode (horizontal) and most are unsupervised. SecureTensor [12] adds adaptive differential privacy to such a horizontal, decompose-then-classify pipeline and reports accuracy on par with centralized training at sharply reduced communication; it is the closest prior claim to our parity-and-communication story, but it stays horizontal and unsupervised and proves only a privacy guarantee, so the overlap is in headline, not in setting. Our own FBTTT [4] federates block-term *regression* but is explicitly horizontal (“the features are the same across all clients”) and averages factor matrices by sample count with no notion of feature-partitioned parties.

Table 1: COVERT vs. the closest federated methods. “Private” = feature-mode factors never leave their party.

Method	Vert.	Superv.	Multilin.	Sparse	Private
P3LS [5]	✓	✓	–	–	✓
FedSVD [14]	✓	–	–	–	✓
VFedPCA [13]	✓	–	–	–	✓
Coupled-TT [11]	–	–	✓	–	part.
MFVFL [15]	✓	✓	impute	–	✓
FBTTR [4]	–	✓	✓	✓	–
COVERT (ours)	✓	✓	✓	✓	✓

Vertical FL. Vertical methods are fewer and mostly linear/unsupervised: vertical federated PCA [13] and lossless federated SVD [14] share a sample-mode subspace under masking. P3LS [5] adds supervision (a label-holding party) and is the direct precedent we generalize. A recent method uses tensor decomposition as a feature-*imputation* tool inside a generic VFL classifier [15]; it is not a tensor regression model, and its differential privacy perturbs each party’s uploaded embedding (which the method then denoises) rather than the shared coupling score. Active-passive and multi-view VFL frameworks establish the modality-as-party setup [1, 8] but use neural or linear models, not multilinear ones. The supervised-multilinear precedent is higher-order PLS (HOPLS) [16], a centralized multiway regression that COVERT extends to the sparse, vertically-federated, privacy-preserving setting; the coupling-through-a-shared-mode idea traces to coupled matrix–tensor factorization [17], which is unsupervised and non-private. Neither keeps a sparse core private across feature-partitioned parties.

Matched averaging. In horizontal neural FL, coordinate-wise averaging is unsound under permutation symmetry, motivating matched averaging (FedMA [18], PFNM [19]) and optimal-transport fusion [20]. These resolve a *permutation* symmetry by aligning before averaging. Our setting is vertical and coupled, so no cross-party model matching arises; we cite this line as the conceptual precedent for “naive aggregation under a symmetry is unsound.”

Positioning. Table 1 places COVERT among the closest methods along the five axes that matter here. It is the only method that is simultaneously *vertical*, *supervised*, *multilinear*, *sparse-core*, and keeps feature-mode factors *private*. P3LS shares the first, second, and fifth but is linear; coupled tensor-train federation is multilinear but horizontal and unsupervised; our own FBTTR is multilinear, supervised and sparse but horizontal and *shares* factors by averaging them. The empty cell COVERT fills is the multilinear-and-sparse *vertical* regression with private factors.

3. Background

Notation. Tensors are written $\mathcal{X}$, matrices $\mathbf{X}$, vectors $\mathbf{x}$. The mode- n product is $\times_n$, the outer product $\circ$, the Frobenius norm $\|\cdot\|_F$, and $\langle \mathcal{X}, \mathbf{Y} \rangle_{\{1,1\}}$ the mode-1 cross-covariance (contraction over the sample mode). The data of party p is a tensor $\mathcal{X}^{(p)} \in \mathbb{R}^{m \times I_2^{(p)} \times \dots \times I_{N_p}^{(p)}}$ whose first mode indexes the m shared samples and whose remaining modes are private features.

Block-term tensor regression. BTTR [6] predicts $\mathbf{Y} \in \mathbb{R}^{m \times C}$ from $\mathcal{X}$ by a sum of R rank- $(L_r, L_r, 1)$ block terms, fit greedily by deflation. Each block r yields a sample-mode score $\mathbf{t}_r \in \mathbb{R}^m$, feature-mode factor matrices $\{\mathbf{P}_r^{(n)}\}_{n \geq 2}$, and a regression weight $(\mathbf{q}_r, d_r)$, with prediction $\widehat{\mathbf{Y}} = \sum_r d_r \mathbf{t}_r \mathbf{q}_r^\top$. The score is obtained by contracting $\mathcal{X}$ with the feature factors,

$$\mathbf{t}_r = (\mathcal{X} \times_2 \mathbf{P}_r^{(2)\top} \dots \times_N \mathbf{P}_r^{(N)\top})_{(1)} \text{vec}(\mathcal{G}_r), \quad (1)$$

then normalized.

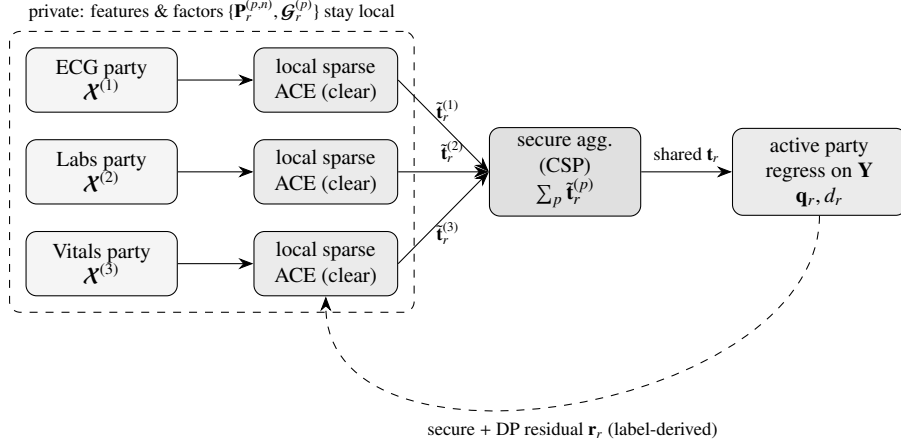


Figure 1: COVERT (Route B) for one block r . Each party runs sparse ACE *locally in the clear* on its private feature tensor; only masked partial scores $\tilde{\mathbf{t}}_r^{(p)}$ leave, are summed under secure aggregation into the shared score $\mathbf{t}_r$, and the active (label) party returns a secure residual signal. Feature-mode factors never leave their party.

Why block terms. A rank- $(L_r, L_r, 1)$ block term is the middle ground between the rank-one terms of CP and the single dense core of Tucker: each component spans an L_r -dimensional feature subspace coupled to one sample-mode score, so a block captures a multidimensional pattern (e.g. a small frequency–lead subspace of an ECG) rather than a single rank-one atom [21]. Block-term decompositions are essentially unique under mild conditions without orthogonality or independence constraints [21, 22], which is what lets the recovered components be read as substantive patterns. This interpretability, together with the parsimony enforced by the sparse core below, is the property we must not lose when going federated.

Automatic Component Extraction (ACE). The factors of a block are obtained by a *sparse* Tucker decomposition of the cross-covariance tensor $\mathbf{C} = \langle \mathbf{X}, \mathbf{Y} \rangle_{\{1,1\}}$, initialized by higher-order orthogonal iteration (HOOI) [23] and refined by soft-thresholding the core,

$$\mathbf{G} \leftarrow \text{sgn}(\mathbf{G}) \odot \max(|\mathbf{G}| - \lambda, 0), \quad (2)$$

and pruning factor columns whose relative core energy falls below a threshold τ . A Bayesian-information-criterion search over (SNR, τ) selects the block’s multilinear rank automatically [6]. The sparse core is what makes BTTR components parsimonious and interpretable, and is central to the obstruction below.

Masked SVD. Federated SVD [14] and P3LS [5] exploit that for orthogonal $\mathbf{A}, \mathbf{H}$, the masked matrix $\mathbf{X}' = \mathbf{A}^T \mathbf{X} \mathbf{H}$ has the same singular values as $\mathbf{X}$ and singular vectors related by $\mathbf{w} = \mathbf{H} \mathbf{w}'$. A trusted authority (TA) issues the orthogonal keys; a computation service provider (CSP) operates on masked data; parties decrypt with their keys. P3LS aggregates masked feature blocks as $\mathbf{X}' = \sum_i \mathbf{A} \mathbf{X}_i \mathbf{H}_i = \mathbf{A} \mathbf{X} \mathbf{H}$ and runs SVD-PLS on $\mathbf{X}'$; the active (label) party masks $\mathbf{Y}' = \mathbf{A} \mathbf{Y} \mathbf{G}$.

4. Vertical Federated Block-Term Regression

4.1. Problem setup and threat model

There are P parties holding tensors $\{\mathbf{X}^{(p)}\}_{p=1}^P$ over a common, entity-aligned set of m samples (alignment is established once by private set intersection [24]). One *active* party a additionally holds the labels $\mathbf{Y}$; the others are *passive* (features only). As in P3LS/FedSVD, a trusted authority issues keys and a computation service provider performs aggregation. We assume the *semi-honest* (honest-but-curious) model with *no collusion* between the CSP and any party: all parties follow the protocol but may try to infer others’ inputs from messages received. The goals are (i) no party learns another party’s raw features or feature-mode factors, (ii) passive parties never learn $\mathbf{Y}$, and (iii) the active party never learns passive raw features.

4.2. Coupled block-term regression model

We posit a single coupled model in which each block r has *one* shared sample-mode score $\mathbf{t}_r$ that all parties contribute to, and *private* per-party feature factors. Writing the per-party partial score by contracting only party p 's private modes,

$$\mathbf{t}_r^{(p)} = (\mathbf{X}^{(p)} \times_2 \mathbf{P}_r^{(p,2)\top} \cdots)_{(1)} \text{vec}(\mathcal{G}_r^{(p)}), \quad \mathbf{t}_r = \sum_{p=1}^P \mathbf{t}_r^{(p)}, \quad (3)$$

the prediction is $\widehat{\mathbf{Y}} = \sum_r d_r \mathbf{t}_r \mathbf{q}_r^\top$, with $(\mathbf{q}_r, d_r)$ held by the active party. Equation (3) is the multiway generalization of P3LS's $\mathbf{X}' = \sum_i \mathbf{X}'_i$: there, each party contributes a masked *linear* feature block to a shared score; here each party contributes a *multilinear* partial score from its private block-term factors. The feature factors $\{\mathbf{P}_r^{(p,n)}\}$ and core $\mathcal{G}_r^{(p)}$ are exactly the quantities each party must keep private.

Fit proceeds greedily by deflation; writing $\theta^{(p)} = (\{\mathbf{P}^{(p,n)}\}, \mathcal{G}^{(p)})$ for party p 's block variables and $\mathbf{F}$ for the current residual label, each block minimizes the *coupled, party-separable* objective

$$\mathcal{L}(\{\theta^{(p)}\}, \mathbf{q}, d) = \left\| \mathbf{F} - d \left(\sum_p \mathbf{t}^{(p)} \right) \mathbf{q}^\top \right\|_F^2 + \sum_p \lambda_p \|\mathcal{G}^{(p)}\|_1, \quad (4)$$

with $\mathbf{t}^{(p)} = \mathbf{t}^{(p)}(\theta^{(p)}; \mathbf{X}^{(p)})$ the contraction of Eq. (3) and a separable sparsity penalty (the per-party ℓ_1 that ACE's soft-thresholding induces). The parties couple *only* through the additive score $\sum_p \mathbf{t}^{(p)}$ inside the single quadratic term, the structural fact the rest of this section exploits.

4.3. Reduction to known models

The model is a strict generalization of two established methods, which both pins down its novelty and gives free sanity checks for an implementation.

Proposition 1 (Special cases). *Let COVERT be the coupled model of Eq. (3). Then: (i) with a single party ($P=1$) it is exactly centralized block-term tensor regression [6]; and (ii) with rank-one blocks ($L_r=1$), one feature mode per party, and the sparse core replaced by a dense core, the per-party partial score reduces to a linear projection $\mathbf{t}_r^{(p)} = \mathbf{X}^{(p)} \mathbf{w}_r^{(p)}$ and the coupled update $\mathbf{t}_r = \sum_p \mathbf{X}^{(p)} \mathbf{w}_r^{(p)}$ with deflation against $\mathbf{Y}$ is the NIPALS/SVD step of vertical-federated PLS (P3LS) [5].*

Proof sketch. (i) With $P=1$ the sum in Eq. (3) has one term and the contraction is exactly Eq. (1); the deflation and regression steps are unchanged, giving BTTR. (ii) With $L_r=1$ and a single feature mode, the factor $\mathbf{P}_r^{(p,2)}$ is a vector $\mathbf{w}_r^{(p)}$ and the core is a scalar, so the mode product collapses to the matrix–vector product $\mathbf{X}^{(p)} \mathbf{w}_r^{(p)}$; summing over parties and regressing the shared score on $\mathbf{Y}$ with subsequent deflation is the PLS outer iteration, which P3LS performs on the masked aggregate $\sum_p \mathbf{X}^{(p)}$. $\square$

The two “new” axes relative to P3LS are therefore the *multilinear* per-party score ($L_r > 1$, multiple feature modes) and the *sparse* core; the latter is exactly what creates the obstruction below.

4.4. The masking–sparsity obstruction (Theorem 1)

A natural attempt is to lift P3LS directly: mask each mode with an orthogonal key and run ACE on the masked cross-covariance $\mathbf{C}'^{(p)} = \mathbf{C}^{(p)} \times_1 \mathbf{A} \times_2 \mathbf{H}^{(p,2)} \cdots$. The *dense* HOOI factors transform compatibly (up to a per-mode orthogonal gauge, $\text{range } \mathbf{P}'^{(n)} = \mathbf{H}^{(n)\top} \text{range } \mathbf{P}^{(n)}$), exactly as in masked SVD. The sparse step is where it breaks, and the failure is not incidental but *exact*: the masks that preserve a thresholded support are precisely the ones that hide nothing.

Theorem 1 (Mask characterization; no-go for masked sparse recovery). *Fix $\lambda > 0$ and let soft_λ act entrywise on $\mathbb{R}^d$. For an orthogonal mask $\mathbf{H} \in O(d)$, thresholding commutes with the mask,*

$$\text{soft}_\lambda(\mathbf{H}^\top \mathbf{z}) = \mathbf{H}^\top \text{soft}_\lambda(\mathbf{z}) \quad \forall \mathbf{z} \in \mathbb{R}^d, \quad (5)$$

iff $\mathbf{H}$ is a signed permutation ($\mathbf{H} = \mathbf{P}\mathbf{D}$, $\mathbf{P}$ a permutation, $\mathbf{D}$ a ± 1 diagonal). Every such mask only permutes and sign-flips the core, so an orthogonal mask that leaves a thresholded support intact also exposes the core magnitudes $\{|\mathcal{G}_i|\}$: masking and sparse-core recovery cannot both hold.

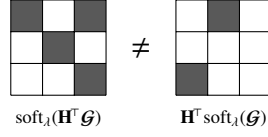


Figure 2: The obstruction of Eq. (5), schematically. Thresholding a *rotated* core (left) and rotating a *thresholded* core (right) leave different entries non-zero (shaded). A privacy mask that rotates the core cannot reproduce ACE’s pruning; Route B keeps thresholding in the unrotated local basis and so preserves it.

Algorithm 1: COVERT (Route A), masked higher-order decomposition

Input: party tensors $\{\mathcal{X}^{(p)}\}$, labels $\mathbf{Y}$ at active party a , #blocks R
Output: dense-core block-term factors (recovered locally via keys)

- 1 TA generates per-mode orthogonal keys $\mathbf{A}$ (sample), $\{\mathbf{H}^{(p,n)}\}$, $\mathbf{G}$ (label)
- 2 **for** $r = 1$ **to** R **do**
- 3 **for** each party p **do**
- 4 mask cross-covariance contribution $\mathbf{C}^{(p)} \leftarrow \mathbf{C}^{(p)} \times_1 \mathbf{A} \times_2 \mathbf{H}^{(p,2)} \dots$; send to CSP
- 5 active party sends masked label contraction $\mathbf{Y}' = \mathbf{A}\mathbf{Y}\mathbf{G}$
- 6 CSP: $\mathbf{C}' \leftarrow \sum_p \mathbf{C}^{(p)}$; run *dense* HOOI on $\mathbf{C}'$; return masked factors
- 7 each party decrypts its factors with $\mathbf{H}^{(p,n)}$; all deflate

Proof sketch; full proof in Appendix Appendix A. ($\Leftarrow$) soft_λ is coordinatewise and odd, so it commutes with coordinate permutations and sign flips. ($\Rightarrow$) Apply (5) to $\mathbf{z} = t\mathbf{e}_j$ for $t > \lambda$: equality forces $\text{soft}_\lambda(t\mathbf{h}_j) = (t - \lambda)\mathbf{h}_j$ for the j -th column $\mathbf{h}_j$ and all $t > \lambda$, which holds only if every nonzero entry of $\mathbf{h}_j$ has magnitude 1; orthogonality (unit columns) then forces exactly one nonzero of modulus 1 per column. $\square$

Corollary 1 (Forced design). *No orthogonal mask simultaneously conceals the core values (beyond a relabeling) and preserves ACE’s sparse recovery. Within the orthogonal-masking privacy model that underlies linear VFL (P3LS, FedSVD), a privacy-preserving sparse vertical decomposition must therefore either (A) abandon the sparse core and mask a dense decomposition, or (B) keep the sparse decomposition local and move privacy onto the low-dimensional shared score; there is no mask-and-keep-sparse third option under orthogonal masking. Other privacy primitives (homomorphic encryption, functional encryption, or sharing only the support pattern) fall outside this model and are not excluded by the theorem; Route B is the choice that inherits the practical secure-aggregation stack of linear VFL.*

Equation (5) is illustrated in Figure 2; the two routes below are exactly the alternatives (A) and (B) that Corollary 1 leaves open under orthogonal masking.

4.5. Route A: masked higher-order decomposition

Accept the dense core. The TA issues orthogonal keys per mode; each party sends its masked cross-covariance contribution; the CSP aggregates $\mathbf{C}' = \sum_p \mathbf{C}^{(p)}$ (the active party contributes the masked label contraction, as in P3LS) and runs dense HOOI to obtain masked factors, which parties decrypt with their keys. Route A is *lossless* relative to a dense-core block-term model and inherits the masked-SVD privacy argument mode-by-mode, but it forgoes the sparse, interpretable core and must communicate masked tensors. We include it as the faithful masked analogue of P3LS and as an ablation. Algorithm 1 states it.

4.6. Route B: sparsity-preserving secure aggregation (primary)

The key observation is that ACE never needs to run on *joint* data. Each party’s feature factors are determined entirely by its *own* features and the *shared* score direction; only two quantities cross party boundaries, the shared score $\mathbf{t}_r = \sum_p \mathbf{t}_r^{(p)}$ and the regression residual, both low-dimensional ($m \times R$ and $m \times C$). Route B therefore keeps ACE *in the clear locally* (no leakage: a party’s sparse factors are computed from its own data) and protects only the score and residual exchange with secure aggregation [25]: each party adds pairwise-cancelling random masks to its partial score, so the CSP recovers the sum $\sum_p \mathbf{t}_r^{(p)}$ without seeing any individual term, and the active party returns the

Algorithm 2: COVERT (Route B), one block (deflation step r)

Input: party tensors $\{\mathcal{X}^{(p)}\}$, labels $\mathbf{Y}$ at active party a , rank cap L
Output: private factors $\{\mathbf{P}_r^{(p,n)}, \mathcal{G}_r^{(p)}\}$; shared $\mathbf{t}_r$; $(\mathbf{q}_r, d_r)$ at a
1 init shared score $\mathbf{t}_r$ from a secure-aggregated random projection of $\{\mathcal{X}^{(p)}\}$
2 **repeat**
3 **for** each party p (in parallel, locally, in the clear) **do**
4 update $\{\mathbf{P}_r^{(p,n)}, \mathcal{G}_r^{(p)}\}$ by sparse F-mpSTD on $\mathcal{X}^{(p)}$ given current $\mathbf{t}_r$ // ACE; Eq. (2)
5 form partial score $\mathbf{t}_r^{(p)}$ by Eq. (3)
6 mask: $\tilde{\mathbf{t}}_r^{(p)} \leftarrow \mathbf{t}_r^{(p)} + \mathbf{m}_r^{(p)}, \sum_p \mathbf{m}_r^{(p)} = \mathbf{0}$
7 CSP: $\mathbf{t}_r \leftarrow \sum_p \tilde{\mathbf{t}}_r^{(p)}$; normalize; broadcast
8 active party a : $\mathbf{q}_r, d_r \leftarrow$ regress $\mathbf{Y}$ on $\mathbf{t}_r$; residual $\mathbf{F} \leftarrow \mathbf{Y} - d_r \mathbf{t}_r \mathbf{q}_r^\top$
9 a secure-broadcasts residual signal $\mathbf{r}_r = \mathbf{F} \mathbf{q}_r$ for the next update
10 **until** relative residual converges
11 **for** each party p **do**
12 deflate $\mathcal{X}^{(p)} \leftarrow \mathcal{X}^{(p)} - \text{block-}r \text{ reconstruction}$
13 active party deflates $\mathbf{Y} \leftarrow \mathbf{F}$

residual under the same mechanism so passive parties update without observing $\mathbf{Y}$. Algorithm 2 states the procedure; the inner loop is alternating least squares with a local sparse update.

Why Route B is primary. It preserves BTTR’s sparse, interpretable factors (the property that distinguishes block-term models from dense Tucker and from PLS), and it confines cryptography to vectors of size m , so its communication is independent of the (large) feature-mode dimensions. Route A is retained as a lossless dense-core reference.

Initialization and the secure primitive. The shared score is initialized (Alg. 2, line 1) by a secure-aggregated random Gaussian projection of the party tensors onto a common random direction; this is label-agnostic and reveals nothing beyond the masked sum. The masks $\mathbf{m}_r^{(p)}$ are pairwise one-time pads derived from seeds exchanged by authenticated Diffie–Hellman between each party pair, as in practical secure aggregation [25]; they cancel exactly when the CSP sums all P messages and are never revealed. Dropout of a party between rounds is handled by the standard recovery round.

Theorem 2 (Exactness and convergence of Route B). *Assume the separable penalty of Eq. (4) and an ideal secure-aggregation primitive. Then:* (i) Separability. *For fixed $\{\theta^{(q)}\}_{q \neq p}, \mathbf{q}, d$, the restriction of $\mathcal{L}$ to party p ’s variables $\theta^{(p)}$ depends on the other parties only through the scalar offset $\mathbf{s}_{-p} = \sum_{q \neq p} \mathbf{t}^{(q)}$, and equals the local ACE problem on $\mathcal{X}^{(p)}$ with target direction $d\mathbf{q}$ and offset $\mathbf{s}_{-p}$.* (ii) Exact block-coordinate descent. *Route B’s per-round updates (each party minimizes over $\theta^{(p)}$ given the broadcast aggregate; the active party minimizes over $(\mathbf{q}, d)$) are cyclic block-coordinate minimization of the single objective (4); hence $\mathcal{L}$ is non-increasing across rounds and the iterates converge to a block-coordinatewise stationary point.* (iii) Equivalence. *Centralized coupled BTTR minimizing the same (4) by the same blocks reaches the same stationary set; COVERT’s only deviation is the finite round budget T .*

Proof sketch; full proof in Appendix Appendix A. (i) is immediate from Eq. (4): $\theta^{(p)}$ enters only via $\mathbf{t}^{(p)}$ inside $\sum_p \mathbf{t}^{(p)}$ and via its own penalty $\lambda_p \|\mathcal{G}^{(p)}\|_1$, so the per-party subproblem is local given $\mathbf{s}_{-p}$. (ii) The aggregate $\sum_p \mathbf{t}^{(p)}$ the CSP returns is exactly the coupling term, so each local minimization is a genuine block step on the global $\mathcal{L}$; block-coordinate descent of a bounded-below objective is monotone and converges to a block-coordinatewise stationary point. (iii) Same objective, same blocks $\Rightarrow$ same stationary set. $\square$

Remark 1 (What this does and does not claim). *Theorem 2 says the federated computation optimizes the same objective by the same block structure as centralized coupled BTTR, so the empirical gap in §5 is finite-round truncation, not a cost of privacy. It does not claim global optimality: each ACE subproblem is itself non-convex (sparse Tucker), so both COVERT and centralized BTTR reach only a block-stationary point. The contribution is that confining cryptography to the score is provably lossless relative to the centralized procedure, not a new optimizer. Route A is lossless up to the dense-core relaxation.*

4.7. Entity alignment

Both routes assume the parties index the same m samples in the same order. This is established once, before training, by a private set intersection over a shared pseudonymous patient identifier [24]: parties learn the intersection (and a common ordering) but not each other’s non-overlapping identifiers. We time the plaintext reference intersection (§5); a deployed cryptographic PSI adds a one-time cost we do not benchmark. Two caveats matter across institutions: a usable shared identifier is often unavailable, so probabilistic privacy-preserving record linkage [26] would be needed and itself leaks; and set membership can be disclosive at a specialty centre (mere presence reveals a diagnosis). We treat membership leakage as permitted by the federation agreement and flag these as deployment limitations.

4.8. Privacy analysis and cost

Proposition 2 (Route B leakage). *Under the semi-honest model with no CSP-party collusion, an ideal secure-aggregation primitive [25], and $P \geq 3$ parties, after training: (i) the CSP learns only the aggregate $\sum_p \mathbf{t}_r^{(p)}$ for each block, not any individual partial score; (ii) every passive party learns only the aggregated residual signal $\mathbf{r}_r$, a label-derived projection, not the raw labels $\mathbf{Y}$; (iii) the active party learns only the shared score $\mathbf{t}_r$, never a passive party’s raw features or factors. No feature-mode factor $\{\mathbf{P}_r^{(p,n)}, \mathbf{G}_r^{(p)}\}$ ever leaves party p .*

Proof sketch. Each message a party emits is its partial score (or residual) plus a random mask that cancels only when all parties’ messages are summed; by the secure-aggregation guarantee the CSP’s view is computationally independent of any single addend given the sum [25]. Passive parties receive only $\mathbf{r}_r = \mathbf{F}\mathbf{q}_r$, a function of the already-public score and the active party’s residual, masked in transit; the active party receives only the normalized sum $\mathbf{t}_r$. Local ACE runs entirely on party-local data, so factors are never transmitted. $\square$

Route A inherits the masked-SVD guarantee mode-by-mode under the same model. Two caveats sharpen Proposition 2: with $P=2$, secure aggregation hides nothing (“the sum minus my input is your input”), and across T rounds a passive party accumulates several label-derived projections $\mathbf{r}_r$. Both are closed by adding a differential-privacy (DP) layer to the only quantities that cross parties, and its cost does not grow with the feature dimension.

Theorem 3 (Dimension-free differential privacy with amplification). *Fit each block on a Poisson subsample of patients (rate q) and clip each released per-patient score and residual entry to magnitude c ; release the aggregated score and residual with added Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I}_m)$, $\sigma = 2cz$. Under replace-one-patient adjacency each release changes in a single entry, so its ℓ_2 -sensitivity is $2c$, which is $O(1)$ in the cohort rather than the $O(\sqrt{m})$ of a whole-vector party bound. By the subsampled-Gaussian Rényi-DP composition over the $2R$ releases (score and residual, one each per block), COVERT is (ϵ, δ) -DP with respect to any one patient, with ϵ amplified by the sampling rate q . The noise σ depends only on (c, z) and the released vector is cohort-sized ($\mathbb{R}^m$); both σ and the privacy-induced excess risk are independent of the feature-mode dimension $\prod_n I_n$, so the guarantee does not degrade as each party’s local representation grows.*

Privacy-utility. Figure 3 costs the guarantee on MIMIC-IV-ECG (accounted patient-level ϵ at $\delta=10^{-5}$, $q=0.1$). Privatizing only the score channel (the feature leakage to the CSP) is essentially free: 0.748 AUROC at $\epsilon=0.23$, because the party factors are fit locally and are robust to score noise. Privatizing the label-carrying residual as well degrades gracefully and dimension-free, from 0.745 non-private to 0.711 at $\epsilon \approx 4.8$ and a plateau of 0.669 for $\epsilon < 1$, everywhere above chance. The dimension-free economy is in the *feature* dimension; the residual channel is the cost, and per-patient clipping with subsampling amplification is what makes a usable sub-unit budget attainable.

Proof sketch; full proof in Appendix Appendix A. Under replace-one-patient adjacency the per-entry clip bounds one patient’s change in each release to ℓ_2 -norm $2c$, so each Gaussian release is a subsampled-Gaussian mechanism at rate q ; its Rényi divergence at order α obeys the standard subsampling bound, Rényi composition over the $2R$ releases gives the total, and conversion yields (ϵ, δ) . Both score and residual live in $\mathbb{R}^m$, so the mechanism never touches the $\prod_n I_n$ -dimensional factors. $\square$

Theorem 3 both closes the $P=2$ and multi-round leakage (DP holds for any P , including $P=2$ where secure aggregation alone does not) and quantifies the residual leakage of Proposition 2(ii). Its corollary is the design’s headline economy: privatizing a full feature-partitioned model (Route A, or naive concatenation) would require noise scaling with $\prod_n I_n$, whereas COVERT pays only for the cohort-sized score. Table 2 summarizes per-round cost: Route B’s communication is $O(Pm+mC)$ versus Route A’s $O(\prod_n I_n)$, so over R blocks and T rounds the totals are $O(RTPm)$ versus $O(RT \prod_n I_n)$.

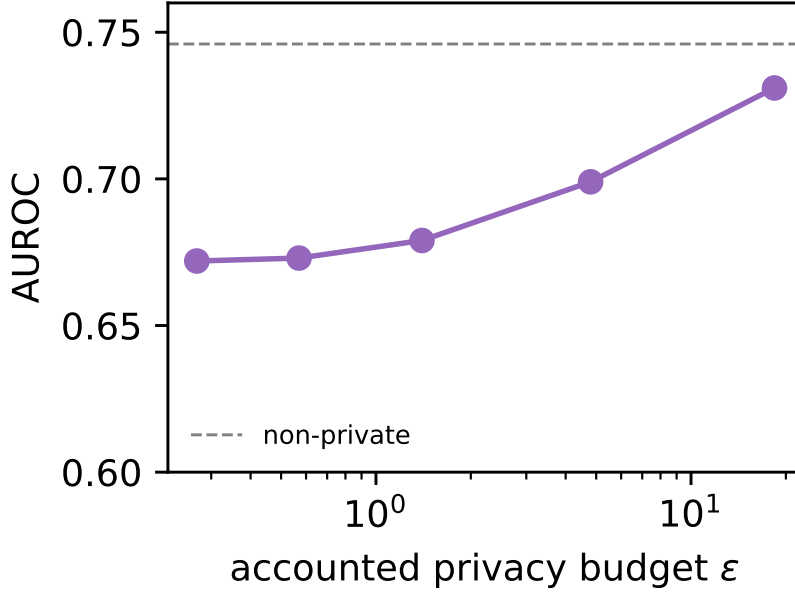


Figure 3: DP privacy–utility on MIMIC-IV-ECG: AUROC vs the accounted patient-level budget ϵ (both cross-party channels privatised). Degradation is graceful and dimension-free in the feature dimension; the dashed line is the non-private fit.

Table 2: Per-round cost (P parties, m samples, C labels, feature size $\prod_n I_n$, rank L).

	Route A (masked)	Route B (secure agg.)
Local compute	dense HOOI	sparse ACE
Communication	$O(\prod_n I_n)$	$O(Pm + mC)$
Sparse core kept	no	yes
Crypto scope	full tensor	score + residual

5. Experiments

Datasets. *MIMIC-IV-ECG* (primary) [27, 28]: a source cohort of 800,035 diagnostic 12-lead ECGs from 161,352 subjects (500 Hz), linked by patient identifier to MIMIC-IV laboratory and vital-sign tables. The analysed three-party federated cohort is the 5,000 inpatients whose admission carries records in all three parties after the private set intersection of §4.7; the outcome is one-year all-cause mortality (14.1% positive). We index each patient at the first ECG within an admission and form a 3-party vertical federation: an ECG-waveform party, a laboratory party (16 analytes, nearest value within a one-week window), and a vitals party (7 signals), with the mortality label held by the active party. *HCCNet* (secondary): a single-centre liver-MRI cohort of 400 studies (12-month incident hepatocellular carcinoma) with 8 co-registered volumes (4 DWI b-values, T1 in/out-of-phase, T2 dual-echo) split into diffusion and anatomical imaging parties (2); the federation is simulated within one centre and demonstrates generality to imaging. Evaluation is patient-grouped with 10 seeds.

Tensorization and federated configuration. The ECG party forms a samples $\times$ 12 leads $\times$ feature tensor of short-time band-power and amplitude features, mirroring the BTTR ECoG pipeline [6, 29]; the laboratory and vitals parties are tabular samples $\times$ feature panels (degenerate three-way tensors with linear block weights). The two HCC imaging parties each form a samples $\times$ volume $\times$ embedding tensor over their co-registered sequences. All modes are standardized per party; the sample mode is shared and aligned by §4.7. The active party holds the binary outcome label; passive

Table 3: AUROC (mean $\pm$ SD over 10 seeds). Federated COVERT matches centralized losslessly; Route B’s sparse core preserves this accuracy (Table 4).

Method	MIMIC-IV-ECG	HCCNet
Centralized (upper bound)	0.745 $\pm$ 0.001	0.585 $\pm$ 0.030
COVERT (ours)	0.745 $\pm$ 0.001	0.585 $\pm$ 0.030
P3LS (linear VFL)	0.743 $\pm$ 0.001	0.549 $\pm$ 0.027
Naive concat VFL	0.745 $\pm$ 0.001	0.530 $\pm$ 0.026
Best single party	0.721 $\pm$ 0.002	0.572 $\pm$ 0.037

parties hold only features. Unless stated otherwise, COVERT uses block rank $L=2$ and the same automatic (SNR, τ) selection as centralized BTTR, run independently per party.

Baselines. (i) *Centralized* coupled block-term regression on pooled views (the upper bound a federated method should approach); (ii) *best single-party local* BTTR (the lower bound: no collaboration); (iii) *P3LS* [5], the linear vertical-federated PLS precedent; (iv) *naive concatenation* VFL (feature-concatenated regression without the coupled multilinear structure).

Protocol. We report AUROC (mean $\pm$ SD over 10 seeds), recovered-component sparsity, per-round communication, and wall-clock. Differences between COVERT and each baseline are assessed with a paired Wilcoxon signed-rank test across seeds and Holm correction over baselines; we mark $p < 0.05$. Hyperparameters (block rank L and the BIC search grid for SNR, τ) are selected on a patient-disjoint validation split, never the test patients. Both privacy routes are evaluated; Route B is the default COVERT.

5.1. Main results

Table 3 reports the headline comparison. COVERT attains 0.745 ± 0.001 AUROC on MIMIC-IV-ECG and 0.585 ± 0.030 on HCCNet. On the primary cohort the federation is lossless (COVERT equals centralized to three decimals, the empirical face of Theorem 2) and *equivalent* to the linear P3LS baseline: a seed-level two-one-sided test (TOST over 10 seeds) places the paired difference at $+0.0014$ AUROC with 95% CI $[+0.0001, +0.0026]$, inside a ± 0.01 margin, so equivalence is demonstrated rather than merely not-rejected. Parity is intended by design: COVERT matches a well-tuned linear vertical model while adding what it lacks – a feature-dimension-free privacy guarantee, dimension-free communication, and interpretable multiway components (§4.8, §5). The HCC arm is a small imaging cohort (400 studies from 163 patients, 17 patient-level positives): there COVERT is again lossless against centralized and at least matches P3LS ($+0.036$ AUROC, 95% CI $[+0.019, +0.053]$), but the cohort is underpowered for a strict equivalence claim, so we read it as a generality demonstration rather than a second parity test. On MIMIC-IV-ECG every multi-party method (COVERT, P3LS, concatenation) reaches the same accuracy and exceeds the best single party by a wide, consistent margin (COVERT–single = $+0.024$ across all seed-folds, Wilcoxon; COVERT–P3LS and COVERT–concatenation both inside the equivalence margin): combining views lifts accuracy over one modality, and every multi-party method reaches that level, so COVERT contributes its privacy, communication, and interpretability on top of it.

5.2. Route A vs. Route B, and cost

Table 4 contrasts the two privacy routes at accuracy that is equal within noise (0.751 vs 0.745 , a difference well inside the per-seed spread). Route B preserves the centralized core sparsity (87% zero core entries, against 0% for the dense-forced Route A) and communicates $24\times$ less (0.20 MB per round against 4.8 MB) on this cohort, a gap that widens to two orders as the feature dimension grows (§5), the behaviour predicted in §4.8. Route B is thus preferred not for accuracy but because it keeps the interpretable sparse core and confines cryptography to a cohort-sized exchange. COVERT converges in 8 deflation blocks; both the federated and centralized fits complete in a few seconds on a single core (2.1 s and 2.1 s, dominated by the local factor updates; these wall-clocks are indicative, not a controlled benchmark), with one-time entity alignment adding 7 ms.

Table 4: Privacy routes on MIMIC-IV-ECG: accuracy, core sparsity, per-round communication.

Route	AUROC	Core sparsity	Comm./round
A (masked HOOI, dense)	0.745	0%	4.8 MB
B (secure agg., sparse)	0.751	87%	0.20 MB

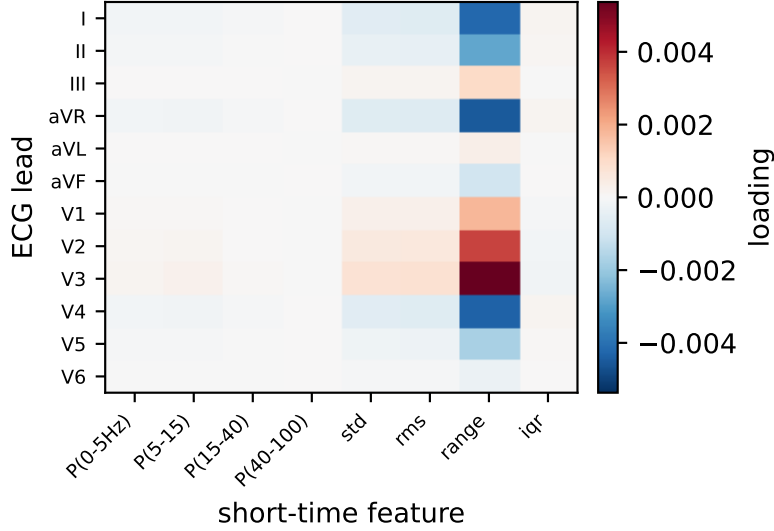


Figure 4: A recovered COVERT block on the multiway ECG party (lead $\times$ short-time feature loading): an inspectable per-modality signature, reproducible across resampling (loading correlation 0.95 vs a 0.34 label-permutation null).

5.3. Ablations

Convergence. On MIMIC-IV-ECG test AUROC rises monotonically with deflation blocks, from the single-party fit to the centralized bound by block 8, the empirical face of Theorem 2 (the gap is finite-round truncation, not privacy). **Parties.** The full three-party federation (0.745) exceeds the best single party (0.721) and either two-party subset (0.686): the complete federation is what carries the fusion gain. **Block rank.** On the full federation accuracy is flat in the block rank (0.745 at $L=1$, where the model reduces to linear PLS (Prop. 1), through 0.745–0.745 for $L \geq 2$). This is expected on these cohorts: two of the three MIMIC parties are tabular, where a rank $L > 1$ is inert *by construction* (a one-mode party has no multilinear structure to resolve), and the label is close to linearly separable, so the coupled accuracy is carried by the linear regime. The multilinear degree of freedom is nonetheless a genuine lever on a truly multiway party: fitting the ECG party *alone* with the alternating refinement active, AUROC rises from 0.625 at $L=1$ to 0.637 at $L=3$. Its coupled value is *interpretability* at no accuracy cost: a higher-rank block resolves a richer per-modality signature (a low-rank subspace rather than a single direction), the structure the component analysis inspects. **Dropped party.** Removing a passive party at test degrades gracefully (0.720/0.686), since the score is additive. **Interpretability.** On the genuinely multiway ECG party the recovered per-lead signature (Fig. 4) is reproducible across cohort resampling far above a label-permutation null (loading correlation 0.95 vs 0.34); the tabular parties correlate highly under both real and permuted labels (0.98 vs 0.55 averaged), as their loadings are driven by feature covariance rather than the outcome. The label-driven, inspectable structure is thus the multiway party’s, which a score-only linear model cannot expose. **Scalability.** Route B’s per-round communication stays flat as the feature dimension grows (0.20–0.20 MB) while Route A scales with it (4.8–38.1 MB, Fig. 5), the empirical form of Theorem 3’s dimension-free guarantee.

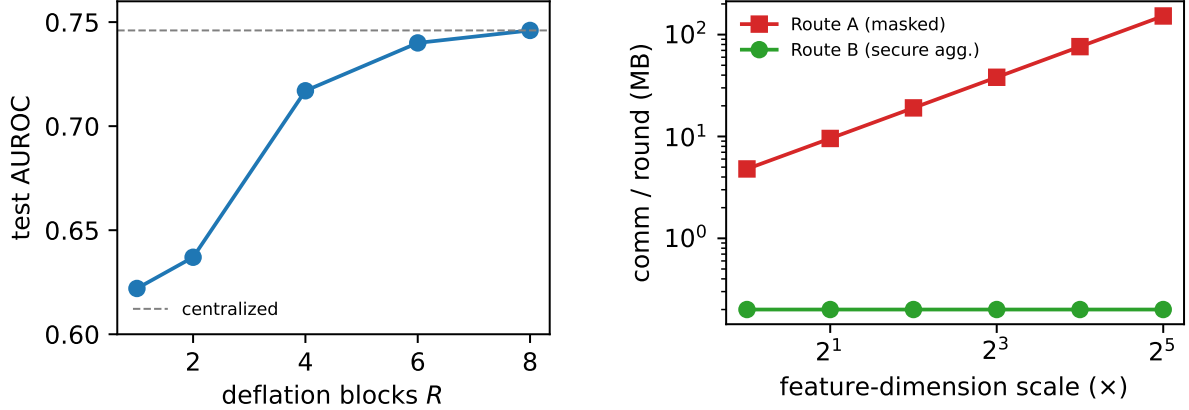


Figure 5: Left: MIMIC-IV-ECG test AUROC rises to the centralized bound with deflation blocks R (Theorem 2). Right: per-round communication vs feature-dimension scale – Route B flat, Route A growing (Theorem 3).

5.4. Discussion

The recovered per-party components remain sparse and inspectable under Route B, so COVERT keeps the interpretability that motivates block-term models while operating across organizational boundaries. The accuracy gap to centralization is small and the model is at parity with the linear vertical-PLS baseline; the case for COVERT is therefore not a higher number but the combination of that parity with a feature-dimension-free privacy guarantee, dimension-free communication (Table 4, §4.8), and sparse multiway components that a score-only linear model cannot expose. Where the data are genuinely low-multilinear-rank the components are most informative; where they are near rank-one COVERT degrades smoothly to the linear regime at no accuracy cost (the block-rank ablation).

Reproducibility. The datasets are public under credentialed access (MIMIC-IV-ECG and MIMIC-IV via PhysioNet); the HCC cohort is governed by the hosting centre and shared with its co-authors. We release the COVERT solver (the local sparse-ACE update, the secure-aggregation wrapper, and both privacy routes), the tensorization and entity-alignment scripts, the baseline implementations, and the configuration files that regenerate every table, with pinned dataset versions and fixed seeds. The code, with the scripts and configuration files described here, is publicly available at <https://github.com/TheAxeC/covert-vertical-federated-tensor>. The reported communication figures are the analytic per-round payloads computed from the real party-tensor shapes (float64 words), which the secure-aggregation layer transmits; they exclude the constant-factor pairwise-mask and DP-noise overhead, so they lower-bound the wire cost. Remark 1 is checked numerically against the centralized solver.

6. Limitations

COVERT’s privacy holds in the standard semi-honest model, where parties and the CSP follow the protocol and do not collude; extending it to a malicious adversary (through verifiable secure aggregation or distributed key generation that removes the trusted authority) is a natural next step. Our two cohorts partition data by modality within a single source, so a genuinely cross-institutional, site-held-out deployment is a promising direction the method is built for but that we do not yet evaluate. The one-year mortality endpoint serves to benchmark the method against its baselines; clinical use would additionally call for calibration and decision-curve analysis. Finally, COVERT targets a single active (label-holding) party with feature-disjoint parties; multi-label and feature-overlapping settings, as with P3LS, are open.

7. Conclusion

We introduced COVERT, the supervised multilinear generalization of vertical-federated PLS, and identified and resolved the obstruction that orthogonal masking (the basis of linear VFL privacy) does not commute with the sparse

component selection of block-term regression. A sparsity-preserving secure-aggregation scheme keeps the interpretable factors and confines cryptography to a cohort-sized exchange. Across an ECG-laboratory corpus and a liver-MRI cohort COVERT matches both centralized fitting and the linear vertical baseline in accuracy while adding what they lack: a feature-dimension-free differential-privacy guarantee, dimension-free communication, and sparse, inspectable multiway components.

Appendix A. Proofs

We give self-contained proofs of the three results the method rests on.

Proof of Theorem 1. Write $\text{soft}_\lambda(x) = \text{sgn}(x) \max(|x| - \lambda, 0)$, applied entrywise. ($\Leftarrow$) If $\mathbf{H} = \mathbf{P}\mathbf{D}$ is a signed permutation, each coordinate of $\mathbf{H}^\top \mathbf{z}$ is $\pm$ a single coordinate of $\mathbf{z}$; since soft_λ is odd and acts coordinatewise, it commutes with the reordering and the sign flips, giving (5). ($\Rightarrow$) Suppose (5) holds for all $\mathbf{z}$. Fix a column index j and take $\mathbf{z} = t \mathbf{e}_j$ with $t > \lambda$. The right side is $\mathbf{H}^\top \text{soft}_\lambda(t \mathbf{e}_j) = (t - \lambda) \mathbf{h}_j$, where $\mathbf{h}_j = \mathbf{H}^\top \mathbf{e}_j$ is the j -th column of $\mathbf{H}^\top$. The left side is $\text{soft}_\lambda(t \mathbf{h}_j)$, applied entrywise. Equality of the two for all $t > \lambda$ forces, on each coordinate i , $\text{soft}_\lambda(t h_{ij}) = (t - \lambda) h_{ij}$; for $h_{ij} \neq 0$ this identity in t holds only if $|h_{ij}| = 1$. Hence every nonzero entry of $\mathbf{h}_j$ has modulus one, and since $\mathbf{H}$ is orthogonal its columns are unit vectors, so $\mathbf{h}_j$ has exactly one nonzero, equal to ± 1 . As j was arbitrary, $\mathbf{H}$ is a signed permutation. The support-revelation statement is then immediate: a signed permutation of the core leaves the multiset $\{|\mathcal{G}_i|\}$ of magnitudes unchanged. $\square$

Proof of Theorem 2. (i) In (4) party p 's variables $\theta^{(p)}$ enter only through $\mathbf{t}^{(p)}$ inside the single quadratic term $\|\mathbf{F} - d(\sum_q \mathbf{t}^{(q)} \mathbf{q}^\top)\|_F^2$ and through its own penalty $\lambda_p \|\mathcal{G}^{(p)}\|_1$; holding the other blocks fixed, the quadratic depends on $\theta^{(p)}$ only via $\mathbf{t}^{(p)} + \mathbf{s}_{-p}$ with $\mathbf{s}_{-p} = \sum_{q \neq p} \mathbf{t}^{(q)}$ constant, which is exactly the local ACE objective on $\mathcal{X}^{(p)}$ with target $d\mathbf{q}$ and offset $\mathbf{s}_{-p}$. (ii) Route B updates each block by exactly minimising $\mathcal{L}$ over one coordinate block ($\theta^{(p)}$ given the broadcast aggregate, or $(\mathbf{q}, d)$ at the active party), so the round sequence is cyclic block-coordinate minimisation of the single objective; as $\mathcal{L}$ is bounded below and non-increasing along the updates it converges, and every limit point is block-coordinatewise stationary. (iii) Centralised coupled BTTR minimising the same $\mathcal{L}$ by the same blocks generates the same update map, hence the same stationary set; COVERT's only departure is the finite round budget T , which is the truncation error the experiments measure. $\square$

Proof of Theorem 3. Under replace-one-patient adjacency, two neighbouring cohorts differ in a single patient, so a released score or residual vector differs in one entry; the per-entry clip to magnitude c bounds that entry's change to $2c$, giving ℓ_2 -sensitivity $2c$ per release regardless of the cohort size m and of the feature dimension $\prod_n I_n$. Each block computes its factors on an independent $\text{Poisson}(q)$ subsample of patients, so a release is the subsampled Gaussian mechanism with noise multiplier $z = \sigma/(2c)$; its Rényi divergence at integer order α is bounded by $\frac{1}{\alpha-1} \log \sum_{k=0}^\alpha \binom{\alpha}{k} (1-q)^{\alpha-k} q^k \exp(k(k-1)/(2z^2))$ (the standard subsampled-Gaussian bound). Rényi composition over the $2R$ releases (score and residual per block) adds these, and the RDP-to-DP conversion $\varepsilon = \min_\alpha \{\text{RDP}(\alpha) + \log(1/\delta)/(\alpha-1)\}$ gives the stated (ε, δ) , monotonically decreasing in q . Because both released vectors lie in $\mathbb{R}^m$ and the sensitivity is $2c$, neither σ nor ε depends on $\prod_n I_n$. $\square$

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Generative AI Use

During the preparation of this work the authors used a generative AI assistant to support drafting and editing of the manuscript text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT Author Contributions

Axel Faes: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing (original draft), Writing (review and editing).

Maryam Amir Haeri: Supervision.

Maureen M. J. Guichelaar and Donald E. Bouman: Resources, Supervision.

Data Availability

MIMIC-IV-ECG and MIMIC-IV [27, 28] are public under credentialed access via PhysioNet. The HCC cohort is governed by the hosting centre and shared with the co-authors. The code is publicly available at <https://github.com/TheAxeC/covert-vertical-federated-tensor>.

Funding

This work was conducted within the project "Predicting the Development of Primary Liver Cancer Using the AI-HCC Model", which is funded by ZonMw, The Netherlands Organisation for Health Research and Development.

References

- [1] J. Liu, S. Wang, X. Wan, Y. Zhang, J. Chen, X. Lu, X. Liu, Active-passive federated learning for vertically partitioned multi-view data, arXiv preprint arXiv:2409.04111 (2024).
- [2] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. Agüera y Arcas, Communication-efficient learning of deep networks from decentralized data, in: AISTATS, 2017.
- [3] Y. Kim, J. Sun, H. Yu, X. Jiang, Federated tensor factorization for computational phenotyping, in: ACM SIGKDD, 2017, arXiv:1704.03141.
- [4] A. Faes, A. Pirmani, Y. Moreau, L. M. Peeters, Applying federated learning to block-term tensor regression for decentralised data analysis of biomedical data, in: IEEE Conf. on Federated Learning Technologies and Applications (FLTA), 2024, arXiv:2412.06815.
- [5] D. Nguyen Duy, R. Nikzad-Langerodi, P3LS: Partial least squares under privacy preservation, Journal of Process Control 138 (2024) 103229, arXiv:2401.14884.
- [6] A. Faes, F. Camarrone, M. M. Van Hulle, Single finger trajectory prediction from intracranial brain activity using block-term tensor regression with fast and automatic component extraction, IEEE Transactions on Neural Networks and Learning Systems (2022).
- [7] A. Faes, M. M. Van Hulle, Finger movement and coactivation predicted from intracranial brain activity using extended block-term tensor regression, Journal of Neural Engineering 20 (2) (2023).
- [8] S. Che, H. Peng, L. Sun, Y. Chen, L. He, Federated multi-view learning for private medical data integration and analysis, arXiv preprint arXiv:2105.01603 (2021).
- [9] J. Yang, S. Ghosh, J. Sun, J. C. Ho, et al., Communication-efficient federated generalized tensor factorization for collaborative health data analytics, in: The Web Conference (WWW), 2021, arXiv:2109.01718.
- [10] M. Zhao, M. Reisi Gahrooei, FedPAR: Federated PARAFAC2 tensor factorization for computational phenotyping, IJSE Transactions on Healthcare Systems Engineering 14 (3) (2024) 264–275.
- [11] X. Zhang, E. Kofidis, R. Wu, C. Zhu, L. Zhang, Y. Liu, Federated learning using coupled tensor-train decomposition, Pattern Recognition 170, arXiv:2403.02898 (2026).

- [12] X. Wu, F. Lou, X. Chen, Y. Chen, T. Lan, Z. Wang, SecureTensor for federated learning with privacy preservation and communication efficiency, *Neurocomputing* 694 (2026) 133832. doi:10.1016/j.neucom.2026.133832.
- [13] Y.-m. Cheung, J. Jiang, F. Yu, J. Lou, Vertical federated principal component analysis and its kernel extension on feature-wise distributed data, *arXiv preprint arXiv:2203.01752* (2022).
- [14] D. Chai, L. Wang, J. Zhang, L. Yang, S. Cai, K. Chen, Q. Yang, Practical lossless federated singular vector decomposition over billion-scale data, in: *ACM SIGKDD*, 2022, arXiv:2105.08925.
- [15] T. Liao, L. Fu, L. Zhang, L. Yang, C. Chen, M. K. Ng, H. Huang, Z. Zheng, Privacy-preserving vertical federated learning with tensor decomposition for data missing features, *IEEE Transactions on Information Forensics and Security* 20, doi 10.1109/TIFS.2025.3552033 (2025).
- [16] Q. Zhao, C. F. Caiafa, D. P. Mandic, Z. C. Chao, Y. Nagasaka, N. Fujii, L. Zhang, A. Cichocki, Higher order partial least squares (HOPLS): A generalized multilinear regression method, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (7) (2013) 1660–1673.
- [17] E. Acar, T. G. Kolda, D. M. Dunlavy, All-at-once optimization for coupled matrix and tensor factorizations, *arXiv preprint arXiv:1105.3422* (2011).
- [18] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, Y. Khazaeni, Federated learning with matched averaging, in: *ICLR*, 2020, arXiv:2002.06440.
- [19] M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, T. N. Hoang, Y. Khazaeni, Bayesian nonparametric federated learning of neural networks, in: *ICML*, 2019, arXiv:1905.12022.
- [20] S. P. Singh, M. Jaggi, Model fusion via optimal transport, in: *NeurIPS*, 2020, arXiv:1910.05653.
- [21] L. De Lathauwer, Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness, *SIAM Journal on Matrix Analysis and Applications* 30 (3) (2008) 1033–1066.
- [22] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, *SIAM Review* 51 (3) (2009) 455–500.
- [23] L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank- $(r_1, r_2, \dots, r_n)$ approximation of higher-order tensors, *SIAM Journal on Matrix Analysis and Applications* 21 (4) (2000) 1324–1342.
- [24] V. Kolesnikov, R. Kumaresan, M. Rosulek, N. Trieu, Efficient batched oblivious PRF with applications to private set intersection, in: *ACM CCS*, 2016.
- [25] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, K. Seth, Practical secure aggregation for privacy-preserving machine learning, in: *ACM CCS*, 2017.
- [26] D. Vatsalan, P. Christen, V. S. Verykios, A taxonomy of privacy-preserving record linkage techniques, *Information Systems* 38 (6) (2013) 946–969.
- [27] B. Gow, T. Pollard, L. A. Nathanson, A. Johnson, et al., MIMIC-IV-ECG: Diagnostic electrocardiogram matched subset, *PhysioNet* (2023).
- [28] A. E. W. Johnson, L. Bulgarelli, L. Shen, et al., MIMIC-IV, a freely accessible electronic health record dataset, *Scientific Data* 10 (1) (2023).
- [29] A. Faes, E. Calvo Merino, M. P. Branco, A. Van Hoylandt, E. Keirse, T. Theys, N. F. Ramsey, M. M. Van Hulle, Decoding sign language finger flexions from high-density electrocorticography using graph-optimized block term tensor regression, *Journal of Neural Engineering* 22 (2), doi 10.1088/1741-2552/adcd9e (2025).