

(SplitNN). In SplitNN, each client trains a partial network up to a cut layer and sends the cut-layer outputs to the server, which completes the remaining forward pass. During backpropagation, gradients at the cut layer are sent back to the client so that local model training can continue [7]. This repeated exchange of intermediate results or gradients can increase communication overhead and operational complexity, especially when training requires many iterations or when clients differ in bandwidth, hardware, storage, and network connectivity [6].

This paper addresses this gap with 10PH-DVFL, a two-tier decoupled VFL architecture positioned within the 10P-Health framework and designed for health data distributed across institutions and modalities. As a proof of concept, the paper evaluates whether this decoupled design preserves predictive performance while reducing cross-silo communication, benchmarked against SplitNN and centralized learning on three clinical datasets.

II. RELATED WORK

A. Healthcare VFL Applications

Several studies have applied VFL to healthcare scenarios where different data holders contain complementary features for overlapping patients. PPVFL-SplitNN [8] combines privacy-preserving record linkage and SplitNN-based collaborative training, evaluated on three healthcare prediction tasks. VFedTrans [9] is a vertical federated knowledge transfer framework evaluated on four real-life medical datasets. QAVFL [10] addresses decentralized multimodal glaucoma screening, incorporating homomorphic encryption and differential privacy. Precision at Heart [11] is an Internet of Things (IoT)-based VFL framework for cardiovascular disease risk prediction across heterogeneous healthcare clients.

B. Representation Learning and Decoupled VFL

SS-VFL [12] uses local contrastive self-supervised learning to train party-specific representation networks; SS-VFL-I freezes representations and sends them once to the server for prediction-head training. DVFL [13] separates training into unsupervised representation learning followed by supervised transfer, removing gradient feedback between participants. APC-VFL [14] trains a local autoencoder at each participant and sends aligned-sample representations once, achieving large communication reductions. VFLHLP [15] combines supervised active and self-supervised passive pre-training before downstream VFL on aligned samples.

C. IoT-Oriented FL and Hybrid HFL-VFL

FedHealthFog [16] introduces fog nodes as local aggregators between edge devices and the cloud for healthcare analytics. A communication-efficient hybrid FL framework for e-health [17] addresses convergence challenges of naively combining HFL and VFL.

III. PROBLEM DEFINITION AND DESIGN REQUIREMENTS

A. Distributed Healthcare Learning Setting

The setting considered here involves several healthcare data silos holding complementary information for partially overlapping patient populations, including clinical, genomic/omics, imaging, and sensor-derived data. The data are vertically partitioned across silos, while some silos may also contain horizontally distributed data sources across devices, gateways, or fog nodes.

One silo acts as the active party because it holds the clinical outcome labels required for supervised learning; in the proposed setting, this is typically a healthcare provider. The remaining silos act as passive parties, holding additional patient features without label access. Cross-silo supervised learning is restricted to the aligned cohort, defined as patients represented across all participating silos. Each silo may also contain unaligned local samples that cannot participate directly in cross-silo supervised learning but can support local representation learning.

B. Design Requirements

The architecture is motivated by four requirements: (i) raw patient-level data must remain within the originating silo to satisfy privacy, governance, and regulatory constraints; (ii) repeated cross-silo exchange of intermediate activations and gradients should be minimized because it increases communication overhead and may introduce privacy risks; (iii) dependence on continuous synchronous participation across silos should be reduced, particularly when a silo relies on intermittently connected devices, gateways, or fog nodes; and (iv) each silo should exploit its full local dataset during representation learning rather than only the aligned intersection, thereby avoiding the underuse of valuable local information.

IV. THE PROPOSED 10PH-DVFL ARCHITECTURE

The proposed 10PH Decoupled Vertical Federated Learning (10PH-DVFL) architecture organizes learning into two sequential, structurally independent tiers. The core design principle is the separation of local encoder pre-training from cross-silo supervised fusion, which eliminates training-phase gradient exchange with passive silos and replaces repeated activation exchange with one-time embedding transfer. Fig. 1 presents the full architecture.

A. Tier 1: Intra-Silo Representation Learning

In Tier 1, each data silo independently trains a modality-specific encoder on its full local dataset using a local learning objective. No inter-silo communication occurs during Tier 1. Each silo may utilize all locally available data, including samples from patients not present in other silos (unaligned samples).

The architecture is instantiated here with three representative silo types. The *clinical/electronic health record (EHR) silo* (active party) holds the outcome labels and pre-trains its encoder using a supervised or self-supervised objective. The

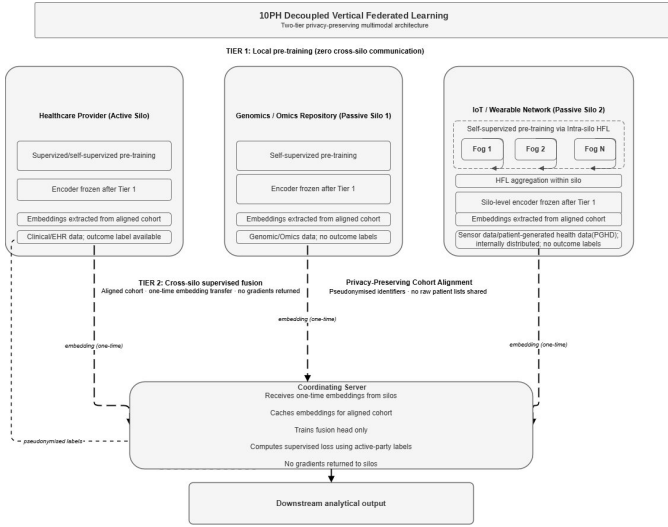


Fig. 1. The proposed 10PH-DVFL architecture. In Tier 1, each silo trains a modality-specific encoder locally using supervised or self-supervised learning. The IoT silo employs intra-silo horizontal FL across fog nodes. After pre-training, all encoders are frozen. In Tier 2, each silo transmits embeddings once to a coordinating server, which caches them, trains only the fusion head using labels from the active party, and returns no gradients to any silo.

genomics/omics silo (passive party) pre-trains its encoder self-supervised on its full local dataset, as no labels are available. For internally distributed IoT or wearable sources, an intra-silo HFL stage lets edge or fog nodes learn a shared silo-level encoder by local training and parameter aggregation at a central IoT hub, with raw sensor data remaining local. Each node then applies the aggregated encoder to its portion of the aligned cohort, and the hub assembles the resulting embeddings into one silo-level matrix transmitted once to the coordinating server. Tier 1 thus yields silo-specific frozen encoders and fixed representations consumed by Tier 2.

B. Tier 2: Cross-Silo Supervised Fusion

Tier 2 performs supervised cross-silo learning over the aligned patient cohort, defined as the set of patients for whom records exist across all participating silos. In this proof-of-concept, aligned records are assumed to have already been identified using pseudonymized identifiers; a production deployment would require a dedicated privacy-preserving entity-resolution mechanism, such as private set intersection (PSI) or privacy-preserving record linkage.

Each silo transmits a single embedding matrix for the aligned cohort to the coordinating server in a pre-fetch transfer. The server caches all embeddings locally and trains only the fusion head, using pseudonymized labels provided by the active party. The supervised loss is back-propagated through the fusion head only. No gradient signals are returned to any silo, and no further communication with passive silos is required after the initial embedding transfer.

C. Communication Cost Analysis

In 10PH-DVFL, Tier 1 incurs zero cross-silo communication, and the total Tier 2 cost is the one-time embedding

payload $C_{DVFL} = N \sum_{s=1}^S d_s b$, where N is the number of aligned samples, d_s the embedding dimension of silo s , S the number of silos, and b the bytes per value. This cost is incurred once, before Tier 2 begins, and is independent of the number of training epochs. SplitNN-based VFL instead transmits cut-layer activations and returns gradients once per batch per epoch, so its communication grows proportionally with training duration.

10PH-DVFL and SplitNN VFL differ architecturally in that 10PH-DVFL exchanges fixed embeddings and labels once over frozen encoders trained on each silo's full local data, whereas SplitNN couples gradients per step over the aligned cohort under synchronous rounds; neither approach provides a formally evaluated privacy guarantee.

D. Architectural Positioning

The proposed design is most closely related to SS-VFL-I [12], which trains local self-supervised representation networks on unlabeled data and sends frozen representations to the server for supervised training with one communication round. 10PH-DVFL extends this through asymmetric pre-training: the active silo may use supervised or self-supervised pre-training, while passive silos rely on self-supervised pre-training. This is related to VFLHLP [15], which also uses supervised active and self-supervised passive pre-training, but unlike VFLHLP, 10PH-DVFL freezes all encoders after Tier 1 and trains only the fusion head. The contribution is therefore not a new privacy primitive or representation-learning objective, but a healthcare-oriented system architecture that combines asymmetric silo-level pre-training, an intra-silo HFL stage, one-time cross-silo embedding transfer, and frozen-encoder fusion, positioned within the 10P-Health framework.

V. EXPERIMENTAL SETUP

A. Datasets and Vertical Partitioning

Three clinical datasets were used. Glioma ($n = 838$ after excluding one incomplete record) [18]: binary grade classification (Low-Grade Glioma [$n = 486$, 58.0%] vs. Glioblastoma Multiforme [$n = 352$, 42.0%]), vertically partitioned into two silos of 13 raw features each. The active silo held demographic variables and ten mutation markers including IDH1, TP53, and ATRX, retaining the labels. The passive silo held the same demographic variables and ten distinct mutation markers including FUBP1, RB1, and NOTCH1, without label access.

Diabetes ($n = 96,146$ after excluding duplicate and inconsistently encoded records) [19]: binary diabetes diagnosis prediction, with 87,664 negative (91.2%) and 8,482 positive (8.8%) cases. The active silo held nine features including body mass index (BMI), glycated haemoglobin (HbA1c) level, blood glucose level, and smoking-history encodings. The passive silo held six features including age, hypertension, heart disease, and gender encoding.

MIDAS ($n = 660$ lesions, 1,980 images after retaining lesions with all three modalities and selecting the first filename-sorted image per modality) [20]: binary skin-lesion

malignancy classification (306 malignant [46.4%], 354 benign [53.6%]) across three imaging modalities defining three vertical silos: dermoscopy as the active, label-holding silo; 6-inch clinical photography and 1-foot contextual photography as passive silos.

B. Model Architectures and Pre-training

For Glioma, each silo used a two-layer multilayer perceptron (MLP) encoder. After one-hot encoding, each silo input contained 27 dimensions. The encoder projected through a 32-unit hidden layer producing a 16-dimensional embedding. The server-side MLP operated on the 32-dimensional concatenated embedding through hidden layers of 16 and 8 units.

For Diabetes, the active and passive silos received 9 and 6 input features respectively. Each encoder produced an 8-dimensional embedding through a 16-unit hidden layer. The server-side fusion head operated on the 16-dimensional concatenation using a linear classification layer.

For MIDAS, each silo used a ResNet50 encoder initialized with ImageNet1K V2 weights, with the final classification layer replaced by an identity mapping to expose 2,048-dimensional feature vectors. In the SplitNN and centralized conditions, each backbone was followed by a trainable projection network that reduced the 2,048-dimensional ResNet50 output to a 256-dimensional modality embedding. In the decoupled conditions, the frozen backbone outputs were reduced to 256 dimensions using Principal Component Analysis (PCA) fitted on the training fold only. In both cases, the fusion head operated on the 768-dimensional concatenation of the three silo embeddings through a 512-unit hidden layer with rectified linear unit (ReLU) activation and dropout of 0.2.

For tabular silos, supervised active pre-training used a binary cross-entropy loss with per-fold positive class weighting; self-supervised pre-training used a Denoising Autoencoder (DAE) [21] with Gaussian noise ($\sigma = 0.1$) and a mean squared error reconstruction loss. For MIDAS, self-supervised pre-training used Bootstrap Your Own Latent (BYOL) [22] with augmentations including random resized cropping, horizontal and vertical flipping, colour jitter, and Gaussian blur.

C. Experimental Conditions and Protocol

For Glioma and Diabetes, two baselines and six 10PH-DVFL configurations were evaluated. Baselines: (1) SplitNN with joint encoder and fusion-head training through repeated activation and gradient exchange; (2) Centralized learning with all features concatenated. The six 10PH-DVFL configurations combined active-silo strategy (SUP: supervised; SSL: self-supervised DAE) with passive-silo strategy (SSL; HFL K10; HFL K20), where K10/K20 denotes the number of independent and identically distributed (IID) simulated intra-silo clients. For MIDAS, four conditions were evaluated: SplitNN, centralized, SSL+SSL, and SUP+SSL.

All experiments used stratified five-fold cross-validation (64% train, 16% validation, 20% test). Fold indices were precomputed and shared across all conditions. All model

TABLE I
TRAINING HYPERPARAMETERS

Setting	Glioma	Diabetes	MIDAS
Optimizer	Adam	Adam	AdamW ^a
Learning rate	10^{-3}	10^{-3}	10^{-4b}
Batch size	64	256 ^c	64
Epochs	100	100	20
Early-stop patience	none	none	7
HFL rounds $\times$ local	100×1	100×1	—
ResNet50 backbone	—	—	trainable ^d

^afusion head: Adam. ^bBYOL: 3×10^{-4} . ^cfusion head: 2048. ^dfrozen with PCA(256) in decoupled conditions. HFL evaluated at $K=10$ and $K=20$; seed 42; all pre-training and fitting performed per training fold.

TABLE II
GLIOMA: MEAN $\pm$ SD ACROSS FIVE FOLDS. BEST 10PH-DVFL IN BOLD.

Method	AUROC	PR-AUC
Centralized	0.9112 ± 0.0277	0.8580 ± 0.0734
SplitNN	0.8937 ± 0.0266	0.8500 ± 0.0420
SSL+SSL	0.8842 ± 0.0245	0.8116 ± 0.0632
SSL+HFL K10	0.8782 ± 0.0294	0.8000 ± 0.0607
SSL+HFL K20	0.8818 ± 0.0226	0.8016 ± 0.0529
SUP+SSL	0.9142 ± 0.0286	0.8560 ± 0.0761
SUP+HFL K10	0.9146 ± 0.0212	0.8620 ± 0.0550
SUP+HFL K20	0.9147 ± 0.0215	0.8622 ± 0.0557

fitting and representation learning, including feature scaling and one-hot encoding, DAE and BYOL self-supervised pre-training, supervised active-silo pre-training, intra-silo HFL aggregation, and PCA fitting, were performed independently within each cross-validation training fold; validation and test samples were excluded from every fitting and representation-learning stage. Primary metrics were the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (PR-AUC), reported as mean $\pm$ sample standard deviation (SD). Experiments ran on Ubuntu 22.04.5 LTS under Windows Subsystem for Linux 2 (WSL2), using Python 3.10, PyTorch 2.5.1, scikit-learn 1.7.2, NumPy 2.2.6, and Flower 1.26.1 [23], with fixed random seed 42. Glioma and Diabetes experiments used CPU only; MIDAS additionally used an NVIDIA GeForce GTX 1650 GPU (4 GB VRAM). Key training hyperparameters are summarized in Table I; complete per-condition configurations, fold indices, and the communication-cost computation are provided in the public code repository (<https://github.com/mahmoud-abuaysheh/10PH-DVFL>).

VI. EMPIRICAL RESULTS AND DISCUSSION

Results are reported as mean $\pm$ SD across five cross-validation folds.

A. Glioma Grade Classification

Table II reports the Glioma results. The three SUP-active configurations exceeded both baselines in mean AUROC; the best, SUP+HFL K20, reached AUROC 0.9147 against SplitNN's 0.8937, a gap within the fold-level SD. The SSL-active configurations were consistently lower (AUROC 0.878–0.884).

TABLE III

DIABETES: MEAN $\pm$ SD ACROSS FIVE FOLDS. BEST 10PH-DVFL IN BOLD.

Method	AUROC	PR-AUC
Centralized	0.9716 $\pm$ 0.0021	0.8514 $\pm$ 0.0124
SplitNN	0.9685 $\pm$ 0.0032	0.8378 $\pm$ 0.0122
SUP+SSL	0.9685 $\pm$ 0.0020	0.8404 $\pm$ 0.0084
SUP+HFL K10	0.9685 $\pm$ 0.0018	0.8386 $\pm$ 0.0075
SUP+HFL K20	0.9684 $\pm$ 0.0017	0.8378 $\pm$ 0.0076
SSL+SSL	0.9507 $\pm$ 0.0025	0.7660 $\pm$ 0.0158
SSL+HFL K10	0.9492 $\pm$ 0.0051	0.7510 $\pm$ 0.0257
SSL+HFL K20	0.9444 $\pm$ 0.0035	0.7257 $\pm$ 0.0177

TABLE IV

MIDAS: MEAN $\pm$ SD ACROSS FIVE FOLDS. BEST RESULTS IN BOLD.

Method	AUROC	PR-AUC
Centralized	0.7530 $\pm$ 0.0570	0.7297 $\pm$ 0.0693
SplitNN	0.7574 $\pm$ 0.0508	0.7088 $\pm$ 0.0656
SSL+SSL	0.7590 $\pm$ 0.0337	0.7340 $\pm$ 0.0598
SUP+SSL	0.7597 $\pm$ 0.0375	0.7188 $\pm$ 0.0603

B. Diabetes Risk Prediction

Table III reports the Diabetes results. The three SUP-active configurations were comparable to SplitNN, with the best (SUP+HFL K10) and SplitNN both at AUROC 0.9685 and centralized slightly higher (0.9716); SSL-active configurations were lower (AUROC 0.944–0.951).

C. MIDAS Skin-Lesion Classification

Table IV reports the MIDAS results. All methods fell within a narrow AUROC band (0.753–0.760), with differences smaller than the fold-level SDs, indicating broadly comparable performance, consistent with cross-validation variability in small-sample settings [24]. The decoupled conditions used frozen ResNet50 features with PCA whereas the SplitNN and centralized conditions used trainable projection networks; the comparison therefore reflects these full pipelines rather than the training protocol in isolation. The complete per-fold results for each experiment are provided in the accompanying GitHub repository.

D. Communication Cost Summary

Table V summarizes the cross-silo communication cost, with 152 $\times$, 169 $\times$, and 21 $\times$ Tier 2 payload reductions for Glioma, Diabetes, and MIDAS; these are cross-silo payload only, not end-to-end system reductions. As analyzed above, this follows directly from one-time embedding caching: unlike SplitNN, where activations and gradients accumulate with training duration, 10PH-DVFL cross-silo communication is bounded by the embedding matrix size and independent of the number of epochs.

E. Interpretation of Findings

The benefit of supervised active pre-training can be explained by the active silo holding the most predictive features. The active silo held the strongest predictors in each task (IDH1 status and age for Glioma; blood glucose, HbA1c, and BMI for Diabetes), letting supervised pre-training optimize directly

TABLE V

CROSS-SILO COMMUNICATION COST COMPARISON.

Dataset	SplitNN	10PH-DVFL	Reduction
Glioma	15.46 MB	0.102 MB	152 $\times$
Diabetes	994.00 MB	5.868 MB	169 $\times$
MIDAS	40.95 MB	$\sim$ 1.93 MB	21 $\times$

toward the outcome. In both cases, SUP-active configurations were comparable to or numerically exceeded SplitNN, while SSL-active configurations underperformed. Because these differences were of the same order as the fold-level SDs, they indicate broadly comparable rather than reliably superior performance. This advantage should be read as conditional on the feature allocation used here: it reflects supervised pre-training performing well when the active silo holds the dominant predictors, and was not tested under alternative partitions that place the strongest predictors at a passive silo.

The underperformance of SSL-active configurations reflects a structural limitation of self-supervised pre-training at the label-holding silo. Trained without supervision, the active encoder may capture general statistical regularities without emphasizing the most discriminative features. Corruption-based reconstruction objectives such as DAE treat all features as equally informative, which misallocates capacity when only a subset carries strong task-relevant signal [25]. This is most consequential precisely when the active silo holds the dominant predictors, as in Glioma and Diabetes. For MIDAS, frozen BYOL-based representations were sufficient for performance comparable to the baselines, indicating that the self-supervised visual objective transferred adequately in this setting.

F. Design Rationale and Generalizability

The decoupling in 10PH-DVFL is a system-level principle: Tier 1 produces a frozen encoder checkpoint at each silo, and Tier 2 trains entirely over cached embeddings, avoiding the instability of naive hybrid HFL-VFL pipelines.

The datasets used here are best understood as evaluating the architectural principle rather than a complete EHR–omics–IoT deployment: Glioma and Diabetes are simulated vertical partitions of tabular datasets, and MIDAS is a naturally multimodal, multi-view imaging dataset, while the clinical/EHR, genomics/omics, and IoT/wearable domains of the 10P-Health framework are represented conceptually rather than through an integrated real-world dataset. In principle the architecture can accommodate additional silos and modalities, provided one silo holds outcome labels, but scalability to larger numbers of institutions and modalities was not evaluated.

G. Limitations

All experiments were conducted as single-machine simulations using Flower’s virtual client engine, enforcing data locality and frozen encoders but not constituting deployment validation in a real multi-institutional environment. The evaluation did not implement privacy-preserving mechanisms such

as differential privacy, homomorphic encryption, or secure aggregation. Transmitted embeddings may also leak information via representation-inversion or inference attacks.

The reported communication reductions concern Tier 2 cross-silo payload only and exclude Tier 1 intra-silo HFL rounds, secure-networking and encryption overhead, entity-alignment cost, protocol metadata, and any subsequent embedding re-transfer after model or data updates. All datasets were fully aligned by construction, not reflecting the partial overlap common in real healthcare environments, where the intersecting cohort may be small and systematically different from each silo's full population, introducing selection bias and precluding the use of missing-modality cases at inference [26]. The frozen-encoder design means changes in tasks or data distributions may require retraining. The HFL component used IID simulated partitions only, without real IoT data, and the evaluation was limited to tabular MLP-based and image ResNet50-based encoders. Evaluation was restricted to discrimination (AUROC and PR-AUC); calibration, clinical utility, and subgroup performance were not assessed.

VII. CONCLUSION

This paper proposed 10PH-DVFL, a two-tier decoupled VFL architecture for communication-efficient multimodal learning within the 10P-Health framework, and evaluated its underlying principle: separating local encoder pre-training from cross-silo supervised fusion, and replacing repeated gradient exchange with one-time embedding transfer over frozen representations. Under the stated assumptions, Tier 2 cross-silo payload communication was reduced by 21- to 169-fold relative to the implemented SplitNN baseline, by design and independently of training duration. Predictive performance was broadly comparable to SplitNN and centralized learning when supervised pre-training was applied at the active silo. 10PH-DVFL is best positioned as a modular, communication-efficient alternative to SplitNN for healthcare settings where data locality, institutional autonomy, and reduced cross-silo synchronization are central design requirements. The experiments do not establish formal privacy guarantees or superiority over existing decoupled VFL approaches. Future work should evaluate privacy attacks and mitigation mechanisms, direct comparison with related decoupled-VFL baselines, alternative feature allocations, partial-alignment scenarios, end-to-end communication and computational cost, and real-world deployment across independent institutional nodes.

REFERENCES

- [1] L. E. Hood and D. J. Galas, "P4 medicine: Personalized, predictive, preventive, participatory," Computing Community Consortium, 2008. [Online]. Available: https://cra.org/ccc/wp-content/uploads/sites/2/2015/05/P4_Medicine.pdf
- [2] N. L. Bragazzi, "From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and augmented medicine," *Patient Preference and Adherence*, pp. 353–359, 2013.
- [3] Health Tech World, "Interview: The future of healthcare and inclusive AI," Jan. 16, 2025. [Online]. Available: <https://www.htworld.co.uk/news/ai/interview-the-future-of-healthcare-and-inclusive-ai>
- [4] Y. Peng, J. Bian, L. Wang, Y. Huang, and J. Xu, "Multimodal federated learning: A survey through the lens of different FL paradigms," 2025, arXiv:2505.21792.
- [5] Z. L. Teo et al., "Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture," *Cell Reports Medicine*, vol. 5, p. 101419, 2024.
- [6] A. Khan, M. ten Thij, and A. Wilbik, "Vertical federated learning: A structured literature review," *Knowledge and Information Systems*, vol. 67, pp. 3205–3243, 2025.
- [7] P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, "Split learning for health: Distributed deep learning without sharing raw patient data," 2018, arXiv:1812.00564.
- [8] B. Alrashed, P. Nanda, H. Dinh, A. Aldahiri, H. Alhosaini, and N. Alghamdi, "PPVFL-SplitNN: Privacy-preserving vertical federated learning with split neural networks for distributed patient data," in *Proc. 22nd Int. Conf. Security and Cryptography (SECRYPT)*, 2025, pp. 13–24.
- [9] C. Huang, L. Wang, and X. Han, "Vertical federated knowledge transfer via representation distillation for healthcare collaboration networks," in *Proc. ACM Web Conference 2023*, 2023, pp. 4188–4199.
- [10] A. Jabbar, J. Huang, M. K. Jabbar, and A. Ali, "Adaptive multimodal fusion in vertical federated learning for decentralized glaucoma screening," *Brain Sciences*, vol. 15, p. 990, 2025.
- [11] S. Shajimon, R. M. Shukla, and A. N. Patra, "Precision at heart: IoT-based vertical federated learning for cardiovascular disease risk prediction," *Journal of the Indian Institute of Science*, vol. 100, pp. 753–772, 2020.
- [12] T. Castiglia, S. Wang, and S. Patterson, "Self-supervised vertical federated learning," in *Workshop on Federated Learning: Recent Advances and New Challenges (NeurIPS 2022)*, 2022.
- [13] S. S. Tripathy, Y. Sirvi, and D. I. Inouye, "Decoupled vertical federated learning for practical training on vertically partitioned data," 2024, arXiv:2403.03871.
- [14] J. Irureta, J. Imaz, A. Lojo, J. Fernandez-Marques, M. González, and I. Perona, "Towards active participant centric vertical federated learning: Some representations may be all you need," 2025, arXiv:2410.17648.
- [15] W. Li, X. Guo, X. Jiao, T. Huang, X. Yan, and Y. Yang, "Vertical federated learning hybrid local pre-training," 2024, arXiv:2405.11884.
- [16] S. S. Tripathy et al., "FedHealthFog: A federated learning-enabled approach towards healthcare analytics over fog computing platform," *Heliyon*, vol. 10, p. e26416, 2024.
- [17] C. Yu, S. Shen, S. Wang, K. Zhang, and H. Zhao, "Communication-efficient hybrid federated learning for E-Health with horizontal and vertical data partitioning," 2024, arXiv:2404.10110.
- [18] K. E. Tasci and K. Camphausen, "Glioma grading clinical and mutation features," UCI Machine Learning Repository, 2022, doi: 10.24432/C5R62J.
- [19] M. Mustafa, "Diabetes prediction dataset," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>
- [20] A. Chiou et al., "MRA-MIDAS: Multimodal image dataset for AI-based skin cancer," Stanford AIMI Center, 2024, doi: 10.71718/15NZ-JV40.
- [21] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Machine Learning (ICML)*, 2008, pp. 1096–1103.
- [22] J.-B. Grill et al., "Bootstrap your own latent: A new approach to self-supervised learning," 2020, arXiv:2006.07733.
- [23] D. J. Beutel et al., "Flower: A friendly federated learning research framework," 2020, arXiv:2007.14390.
- [24] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage*, vol. 180, pp. 68–77, 2018.
- [25] T. Ucar, E. Hajiramezanali, and L. Edwards, "SubTab: Subsetting features of tabular data for self-supervised representation learning," 2021, arXiv:2110.04361.
- [26] J. Sun et al., "Communication-efficient vertical federated learning with limited overlapping samples," 2023, arXiv:2303.16270.