PreComp-VFL: Enhancing Embedding Representation in Vertical Federated Learning for Medical Data

Bashair Alrashed, Priyadarsi Nanda, Dinh Thai Hoang
Faculty of Engineering and Information Technology
University of Technology Sydney, Sydney, Australia
Email: BashairAliM.Alrashed@student.uts.edu.au

Saleh Alqahtani, Raddad Faqihi
College of Computing and Informatics
Saudi Electronic University, Saudi Arabia
Email: sm.alqahtani@seu.edu.sa

Osama Dighriri
College of Engineering and Computer Science
Jazan University, Saudi Arabia
Email: omdighriri@jazanu.edu.sa

Ethar Alzaid, Amani Aldahiri
College of Computer Science and Engineering
University of Jeddah, Jeddah, Saudi Arabia
Email: ezalzaid@uj.edu.sa

***Abstract*—Vertical federated learning (VFL) in healthcare enables institutions holding different features of the same patients to collaboratively train machine learning models without exposing raw data. However, most existing VFL approaches focus on secure aggregation and privacy-preserving computation, while overlooking client-side preprocessing. This limitation can result in less informative embedding representations and increased communication overhead. To address this, we propose PreComp-VFL, a client-level preprocessing method that integrates unsupervised feature selection with dimensionality reduction techniques. PreComp-VFL allows clients to transmit compressed, informative, and privacy-preserving embeddings without requiring label access. We used four real-world medical datasets in a number of experiments demonstrating our proposed scheme and achieved improved model accuracy and F1 Score compared to standard VFL. We also show that our proposed scheme achieves significant reductions in communication costs relative to server-side feature selection scheme.**

***Index Terms*—privacy-preserving, vertical federated learning, embedding representation, feature selection, medical data.**

I. Introduction

The rapid digitization of healthcare has resulted in vast amounts of electronic medical records distributed across hospitals, clinics, and laboratories. These records are often vertically partitioned, meaning that each institution holds a different subset of features describing the same patients. For example, one may store clinical notes, while another manages lab results or genetic markers. Building predictive healthcare models from such fragmented data requires collaboration while strictly preserving patient privacy.

Vertical Federated Learning (VFL) has emerged as a promising solution for privacy-preserving collaboration in healthcare [1]. It is a category of federated learning that allows institutions with disjoint features to train machine learning models collaboratively without sharing raw data. In the VFL-based medical data management framework, each party holds a distinct feature subset for the same group of patients. To support deep learning in this setting, Split Neural Networks (SplitNNs) are widely adopted [2]. SplitNNs divides the global model between clients and a centralized server. This setup allows clients to transmit only intermediate representations, known as embeddings, rather than raw features or full model updates. This architecture not only preserves data privacy, but also reduces communication overhead and adapts well to heterogeneous data distributions.

Despite significant progress in VFL, most existing work in medical data management focuses on model aggregation and privacy-preserving computation [3], [4]. However, less attention has been given to client-side feature preprocessing, especially in the healthcare domain where data is often sensitive, sparse, or high-dimensional [5], [6]. This gap often leads to embeddings that are noisy or redundant, degrading model accuracy and increasing communication overhead. Moreover, certain features known as quasi-identifiers (e.g., age, gender) may pose a re-identification risk even when raw data is not shared. Removing such attributes during preprocessing can enhance privacy protection while also reducing noise and improving model generalization. Therefore, a critical challenge remains unresolved: **how to ensure that client-generated embeddings are both informative and privacy-preserving, particularly when no labels are available at the client side**. Many existing feature selection strategies in VFL rely on supervised feature selection methods or pre-trained models [7], [8]. These strategies are impractical in medical contexts due to ethical or regulatory restrictions, such as HIPAA and GDPR [9]. While some client-local preprocessing studies exist [8], they are computationally intensive and rely on iterative procedures, limiting scalability.

To address these limitations, we propose PreComp-VFL, a client-side preprocessing strategy for VFL in healthcare. PreComp-VFL performs unsupervised local feature selection and compression before collaborative training begins. **Unlike server-based methods that rely on label access, PreComp-VFL enables each client to independently process their data using unsupervised techniques** such as Quasi-Identifier

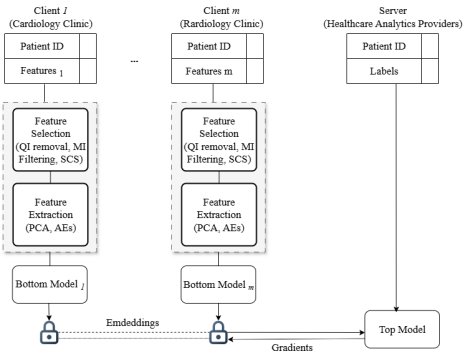

Fig. 1: System Architecture and Data Flow in PreComp-VFL.

(QI) removal, Mutual Information (MI) filtering, and Smart Correlation Selection (SCS). Each client can then combine the selected features with an appropriate dimensionality reduction method, such as Principal Component Analysis (PCA) or AutoEncoders (AEs), depending on the data characteristics and computational resources. Our experiments evaluate the proposed approach across four real-world medical datasets. The results show that PreComp-VFL improves embedding quality, reduces communication cost, and maintains predictive performance under realistic VFL constraints. On average, it reduces communication cost by 20–30% and improves F1 Score by up to 2% compared to standard VFL, without requiring label access. To the best of our knowledge, this is the first work to apply a comprehensive set of unsupervised feature selection and compression techniques locally to enhance embedding representations prior to collaborative training. We summarize the contribution of this paper as follows:

- We propose PreComp-VFL, a client-side preprocessing strategy that enables each participant to transmit compact, informative, and privacy-aware embeddings. This reduces bandwidth usage and enhances the effectiveness of collaborative training without exposing raw features.
- We introduce a modular unsupervised feature optimization pipeline, combining advanced feature selection (QI removal, MI filtering, SCS) with dimensionality reduction (PCA or AEs). This flexibility allows PreComp-VFL to adapt to heterogeneous medical datasets without requiring label access.
- We conduct a comprehensive evaluation on four real-world medical datasets. The results show that PreComp-VFL enhances embedding quality, reduces communication overhead, and maintains strong predictive performance under realistic VFL constraints.

II. System Architecture and Data Flow in PreComp-VFL

PreComp-VFL supports collaborative learning on vertically partitioned medical data while preserving privacy. As shown in Figure 1, the system consists of a central server (active party) holding labels and multiple clients (passive parties), each with disjoint feature subsets. Clients locally preprocess their features using QI removal, MI filtering, or smart correlation-based pruning. Each client applies QI removal, MI filtering, or smart correlation selection to retain informative, non-redundant features without label access, improving generalization and reducing communication. QI removal is used to remove QI features to preserve individual privacy. Although these attributes may not be uniquely identifying on their own, they could potentially re-identify individuals when combined with other attributes. Therefore, clients may choose to eliminate such QIs from their feature space to enhance privacy and potentially improve model generalization. This step is particularly relevant when QIs contribute little to predictive performance. On the other hand, MI is used to capture linear and non-linear dependencies [10], while correlation filtering is used to remove multicollinearity using Pearson coefficients [11]. This is followed by dimensionality reduction. For dimensionality reduction, clients use PCA or AEs. PCA produces linearly uncorrelated components, whereas AEs capture non-linear feature interactions [6]. Both approaches reduce dimensionality and overhead while preserving predictive utility in federated settings. The resulting compressed features are fed into client-side bottom models, whose outputs (embeddings) are transmitted to the server. The server concatenates these embeddings, computes the global loss, and backpropagates gradients to update both global and local model parameters, following the SplitNN paradigm [2]. This workflow reduces communication and computation while mitigating privacy leakage.

Algorithm 1 PreComp-VFL: Local Preprocessing and Compression in VFL

Require: Local features X_m, learning rate η, batch size β, rounds T
Ensure: Compressed features $\hat{X}_m$, Global model Θ

Server: Initialize $\Theta^{(0)}$ and send to clients
for all clients $m = 1 \dots M$ in parallel **do**
  $\hat{X}_m \leftarrow \texttt{PreComp}(X_m)$; feed to bottom model
end for
for each round $t = 0 \dots T-1$ **do**
  for all clients $m = 1 \dots M$ in parallel **do**
    $h_m \leftarrow \sigma(\theta_m \cdot \hat{X}_m)$; send h_m to server
  end for
  Server: Concatenate $\{h_m\}_{m=1}^{M}$, compute loss $L^{(t)}$
  Update $\Theta^{(t+1)} \leftarrow \Theta^{(t)} - \eta \nabla_\Theta L^{(t)}$
  Send $\nabla_{h_m} L^{(t)}$ to clients
  for all clients $m = 1 \dots M$ in parallel **do**
    Update $\theta_m^{(t+1)} \leftarrow \theta_m^{(t)} - \eta \nabla_{\theta_m} L^{(t)}$
  end for
end for

Algorithm 1 outlines the workflow. Each client applies feature selection and dimensionality reduction to transform its private features X_m into compressed representations $\hat{X}_m$, where $k_m \ll d_m$. These are passed through local bottom models to generate embeddings h_m, which are aggregated by the server. The global model parameters Θ are updated using the Adaptive Moment Estimation (ADAM) optimizer, and gradients ∇_{h_m} are returned to clients to update their local parameters θ_m. This iterative process continues until convergence.

III. Performance Evaluation

This section outlines the experimental setup used to evaluate the proposed method. It also provides detailed information on data preprocessing and training procedures.

TABLE I: Evaluation Metrics using All Original Feature (Baseline) + Feature Extraction (PCA, AEs).

Dataset Name	Training Samples	Testing Samples	All Original Features		With PCA		With AEs	
			Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score
Breast Cancer	559	140	97.85	96.77	97.85	97.24	98.57	97.87
Diabetes Prediction	5908	1478	88.70	89.95	88.22	89.42	87.82	88.56
Gliomas	671	168	88.09	86.66	89.88	88.59	87.5	84.67
Chronic Kidney	320	80	98.75	98.55	98.75	98.7	98.75	98.7

TABLE II: Overview of Feature Reduction Rate with/without Feature Extraction.

Database Name	Feature number in		Feature Retained (%) using			
	Client *1*	Client *m*	PCA		AEs	
			Client *1*	Client *m*	Client *1*	Client *m*
Breast Cancer	4	5	25	25	25	25
Diabetes Prediction	7	7	25	25	28	28
Gliomas	14	14	21	14	28	28
Chronic Kidney	13	12	23	33	23	25

A. Datasets

We evaluate the proposed method on four publicly available medical datasets: Breast Cancer [12], Diabetes Prediction [13], Glioma Grading [14], and Chronic Kidney Disease [15]. These datasets were chosen because of their tabular format. To assess robustness, we included datasets of varying sizes and feature dimensions, ranging from small-scale to large-scale collections.

B. Preprocessing and Training Setup

Maintaining data quality is critical for model learning. For the Breast Cancer, Glioma, and Chronic Kidney Disease datasets, missing values were imputed with the feature-wise mean, while the Diabetes Prediction dataset required removal of over 3,800 duplicate records and exclusion of entries with an ambiguous gender or missing smoking history. Duplicates in smaller datasets were retained due to sample scarcity. After cleaning, each dataset was vertically partitioned across two passive clients, with the active party holding labels. We also split each dataset into training/testing sets at an 8:2 ratio. To accommodate dataset complexity, we defined two SplitNN architectures: for Diabetes and Breast Cancer, clients used two fully connected layers (16 and 8 ReLU units) with a server-side single-layer sigmoid head, while Glioma and Chronic Kidney Disease used deeper client models (32 and 16 units) and a three-layer server head (16, 8, 1 units with ReLU and sigmoid). Each client's autoencoder comprised a single encoder–decoder pair, with compression ratios (0.25, 0.5, 0.75) tuned via grid search, where 0.25 consistently balanced dimensionality reduction and performance. Models were initialized using PyTorch defaults and trained with the ADAM optimizer (lr=0.001). Binary cross-entropy was applied for classification tasks, while mean squared error was used for unsupervised AEs training. Each experiment ran for 200 communication rounds, with evaluations performed every 50 epochs.

C. Simulation Results and Analysis

The experimental results obtained from the proposed method are presented. Performance metrics, such as Test accuracy (Acc.) and F1 Score, obtained through the proposed method have been compared against a baseline SplitNN model trained on the full original feature set, with and without feature extraction. Additionally, we conduct a theoretical analysis of training time and computational complexity to compare PreComp-VFL with server-side feature selection and other related methods.

1) Effect of Feature Compression on Model Performance and Dimensionality: This experiment evaluates the impact of compressing original features (baseline) using PCA and AEs on model performance. Table I presents the accuracy and F1 Score of models trained on original features compared to those trained on features compressed using PCA and AEs. Applying feature extraction either slightly improved or maintained model performance while significantly reducing the number of features transmitted per client, as shown in Table II. For PCA, the optimal variance retention threshold was 0.9, selected based on a grid search over thresholds ranging from 0.5 to 0.9. This experiment demonstrates that unsupervised feature extraction effectively reduces communication overhead without compromising model accuracy, particularly when using PCA on small to medium-sized medical datasets. While PCA generally yielded more stable results, AEs underperformed slightly on small datasets such as Glioma, likely due to overfitting. The performance gap is attributed to AEs requiring a larger training sample size and more careful hyperparameter tuning. Future improvements to AE-based compression could include using deeper architectures or introducing regularization techniques to reduce overfitting in low-sample scenarios.

2) Impact of Quasi-Identifier Removal on Privacy and Model Performance: This experiment evaluates how removing QI attributes, such as age and gender, affects both privacy preservation and model performance. It is a common privacy-preserving technique in medical data sharing. We run this experiment only on three datasets: Diabetes Prediction, Gliomas, and Chronic Kidney. We exclude the Breast Cancer Wisconsin dataset because it does not contain any QIs attributes.

Table III reports the performance of the model when QIs are removed prior to training. The results show that for most datasets, removing QIs leads to only marginal decreases in both accuracy and F1 Score, typically less than 2%. For example, in the Diabetes Prediction dataset, accuracy drops from 88.70% to 87.07%, indicating a minimal impact on model performance while improving privacy. However, in some cases, QIs contribute predictive value. For instance, age is a known risk factor for both Glioma and Diabetes. Removing such features may slightly alter the model's discriminative capacity. However, in the Glioma dataset, the F1 Score actually increased marginally from 86.66% to 86.79%, suggesting that QI removal does not always degrade performance. Table III

TABLE III: Performance Metrics Before and After using QI Removal + Feature Extraction.

Dataset Name	Training Samples	Testing Samples	All Original Features		After Drop QI		Drop QI+PCA		Drop QI+AEs	
			Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score
Diabetes Prediction	5908	1478	88.70	89.95	87.07	88.24	86.67	87.84	84.57	85.53
Gliomas	671	168	88.09	86.66	87.5	86.79	88.69	88.19	86.3	86.22
Chronic Kidney	320	80	98.75	98.55	98.75	98.46	98.75	98.63	98.75	98.46

TABLE IV: Performance Metrics Before and After using MI Filter + Feature Extraction.

Dataset Name	Training Samples	Testing Samples	All Original Features		After MI Filter		MI Filter+PCA		MI Filter+AEs	
			Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score
Breast Cancer	559	140	97.85	96.77	97.85	96.9	97.85	97.47	98.57	98.07
Chronic Kidney	320	80	398.75	98.55	98.75	98.41	97.5	97.5	96.25	95.8

TABLE V: Evaluation Metrics Before and After using SCS + Feature Selection.

Dataset Name	Training Samples	Testing Samples	All Original Features		After SCS		SCS+PCA		SCS+AEs	
			Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score	Test Acc.	F1 Score
Breast Cancer	559	140	97.85	96.77	97.14	96.22	98.57	98	97.85	96.77
Gliomas	671	168	88.09	86.66	89.28	86.95	90.47	88.88	87.5	86.79
Chronic Kidney	320	80	98.75	98.55	98.75	98.66	97.5	97.29	95	95.83

also compares the model performance after applying PCA or AEs on top of QI-removed features. Consistent with earlier observations, PCA outperforms AEs across most datasets. This is likely due to the small sample sizes and linear feature interactions, which favor PCA's statistical nature over the higher capacity of AEs. These findings suggest that QI removal is a viable privacy-preserving strategy in medical VFL, particularly when combined with dimensionality reduction techniques like PCA that mitigate any minor loss in model utility.

3) Effect of Mutual Information-Based Feature Selection on Model Performance: This experiment evaluates the impact of MI based feature selection on model performance. We conducted this experiment on the Breast Cancer and Chronic Kidney datasets, excluding Glioma and Diabetes due to a lack of meaningful inter-feature dependency as observed during the threshold tuning process. The MI threshold was selected through a grid search over the range [0.5, 0.6, 0.7, 0.8, and 0.9]. The optimal MI threshold was identified as 0.7, which offered the best balance between removing redundancy and maintaining accuracy.

Table IV presents results after applying MI-based feature selection. The performance metrics remain stable or slightly improved across datasets, indicating that the retained features preserve most of the discriminative information. For example, in the Breast Cancer dataset, the F1 Score improved from 96.77% to 96.9%, even after removing redundant or weakly informative features based on MI. In addition, Table IV combines MI with feature extraction techniques: PCA and AEs. The results show that MI followed by AEs achieved the highest F1 Score (98.07%) for the Breast Cancer dataset, demonstrating the benefit of combining feature selection with dimensionality reduction. These results suggest that MI-based selection is effective for reducing dimensionality without sacrificing accuracy, especially when paired with AEs or PCA.

4) Effect of Smart Correlation Selection-Based Feature Selection on Model Performance: This experiment evaluates the impact of SCS in identifying and removing redundant features using Pearson correlation coefficients. We tested this method on the Breast Cancer Wisconsin, Gliomas, and Chronic Kidney datasets. The Diabetes dataset was excluded due to weak inter-feature correlation. The correlation threshold was selected based on a grid search over the range [0.5, 0.6, 0.7, 0.8, 0.9]. The grid search determined that a correlation threshold of 0.8 provided the best balance between reducing redundancy and maintaining model performance.

Table V presents the results of applying SCS alone. While performance remains comparable on the Breast Cancer and Chronic Kidney datasets, we observe a noticeable improvement in the Gliomas dataset, where accuracy increased from 88.09% to 89.28%. This suggests that SCS is particularly effective in datasets with a high degree of inter-feature correlation. Table V also combines SCS with dimensionality reduction techniques. The combination of SCS and PCA yielded the highest F1 Score of 88.88% on the Gliomas dataset, further supporting the utility of this two-stage preprocessing. These findings highlight the importance of tailoring correlation-based filtering to dataset characteristics. SCS is especially beneficial when feature collinearity is present, as it complements dimensionality reduction techniques like PCA.

D. Runtime and Communication Cost Analysis

In VFL, the overall system efficiency depends heavily on both computational runtime and communication overhead. Traditional server-side feature selection approaches require each client to transmit full-dimensional embeddings to the central server, where global feature selection is then performed. In this study, we denote communication cost as C_{comm} and computational cost as C_{comp}, expressed in terms of sample size n, number of original features per client d, number of clients m, and compressed dimension k. We now contrast PreComp-VFL with server-side selection in terms of theoretical runtime and communication complexity.

In server-side approaches, each client transmits (nd) embedding size per round to the server, resulting in a total communication cost of $(C_{comm} = mnd)$. The server then performs feature selection over the entire feature space of size (md), with methods like Shapley scoring requiring up to $(C_{comp} = (md)^2)$ time due to iterative optimization or pairwise evaluations [16]. This centralized computation becomes a major bottleneck in large-scale or privacy-sensitive applications. In contrast, the proposed PreComp-VFL approach shifts feature selection and compression to each client prior to collaborative training. Each client locally reduces its dimensionality from d to k using unsupervised techniques such as MI, SCS, PCA, or AEs. The computational cost varies by method. For instance, MI and SCS incur $(C_{comp} = d \log d)$ due to pairwise evaluations among features [17], [18], while PCA involves eigenvalue decomposition, requiring $(C_{comp} = d^2 n + d^3)$ [19]. Shallow autoencoders, in contrast, require $(C_{comp} = edn)$ training operations over e epochs [20]. Because the dimensionality is reduced locally, each client transmits only $(C_{comm} = nk)$ compressed embeddings per round, reducing the total communication cost to $(C_{comm} = mnk)$, with $k < d$.

Beyond theoretical complexity, it is important to compare PreComp-VFL with other recent communication-efficient VFL strategies. In particular, we examine the neuron selection approach proposed by [8], which also operates at the client level. However, it differs in when and how compression is applied. The authors perform client-side pruning of the local output neurons based on Taylor approximations. This method requires a warm-up phase with full communication before pruning is applied. It sends a binary mask to the server to indicate the top-d_t neurons to retain. While effective in reducing communication in later rounds, it is architecture-dependent and assumes differentiable gradient behavior. In contrast, PreComp-VFL provides immediate compression, supports non-neural models, and works without label access. This makes it particularly useful in privacy-sensitive healthcare settings. Furthermore, PreComp-VFL avoids the need for raw feature sharing or centralized optimization, reducing both C_{comm} and C_{comp} from the outset. By decentralizing preprocessing, it improves scalability and efficiency, especially in environments with heterogeneous feature spaces and constrained bandwidth.

IV. Conclusion

We proposed PreComp-VFL, a client-side preprocessing strategy for VFL in healthcare. The proposed strategy performs unsupervised feature selection and dimensionality reduction locally, without requiring label access. It improves embedding quality, reduces communication overhead, and enhances data privacy. The method is model-agnostic and well-suited for real-world medical collaborations with heterogeneous data distributions. While effective, PreComp-VFL depends on well-tuned hyperparameters (e.g., PCA thresholds, autoencoder compression) and may face limitations with small datasets or large-scale federations. Autoencoders, in particular, may perform poorly in low-sample settings without proper regularization. Future work will explore integration with secure aggregation protocols to further protect compressed embeddings and evaluate scalability in broader deployments. By addressing these challenges, PreComp-VFL offers a practical path toward scalable, privacy-preserving federated learning for sensitive medical applications.

References

[1] S. Feng, "Vertical federated learning-based feature selection with non-overlapping sample utilization," *Expert Systems with Applications*, vol. 208, p. 118097, 2022.

[2] P. Vepakomma, O. Gupta, T. Swedish, and R. Raskar, "Split learning for health: Distributed deep learning without sharing raw patient data," *arXiv preprint arXiv:1812.00564*, 2018.

[3] X. Qiu, H. Pan, W. Zhao, C. Ma, P. P. B. de Gusmão, and N. D. Lane, "Efficient vertical federated learning with secure aggregation," *arXiv preprint arXiv:2305.11236*, 2023.

[4] H. Sun, Y. Zhang, Z. Xu, R. Zhang, and M. Li, "MK-FLFHNN: A privacy-preserving vertical federated learning framework for heterogeneous neural network via multi-key homomorphic encryption," in *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 552–558, IEEE, 2023.

[5] 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 *In Proceedings of the 22nd International Conference on Security and Cryptography*, pp. 13–24, 2025.

[6] A. Khan, M. ten Thij, and A. Wilbik, "Communication-efficient vertical federated learning," *Algorithms*, vol. 15, no. 8, p. 273, 2022.

[7] T. Castiglia, Y. Zhou, S. Wang, S. Kadhe, N. Baracaldo, and S. Patterson, "LESS-VFL: Communication-efficient feature selection for vertical federated learning," in *International Conference on Machine Learning*, pp. 3757–3781, PMLR, 2023.

[8] C. Yang, Y. Cheng, X. Yang, and Y. He, "A communication-efficient vertical federated learning via neuron selection," in *2023 IEEE Smart World Congress (SWC)*, pp. 1–8, IEEE, 2023.

[9] R. S. Antunes, C. André da Costa, A. Küderle, I. A. Yari, and B. Eskofier, "Federated learning for healthcare: Systematic review and architecture proposal," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 4, pp. 1–23, 2022.

[10] D. Eklund, A. Iacovazzi, H. Wang, A. Pyrgelis, and S. Raza, "BMI: Bounded mutual information for efficient privacy-preserving feature selection," in *European Symposium on Research in Computer Security*, pp. 353–373, Springer, 2024.

[11] H. Zhou, Z. Deng, Y. Xia, and M. Fu, "A new sampling method in particle filter based on pearson correlation coefficient," *Neurocomputing*, vol. 216, pp. 208–215, 2016.

[12] W. Wolberg, "Breast Cancer Wisconsin (Original)." UCI Machine Learning Repository, 1990. DOI: https://doi.org/10.24432/C5HP4Z.

[13] M. Mustafa, "Diabetes prediction dataset." https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset, 2023.

[14] Tasci, Erdal, Camphausen, Kevin, Krauze, Andra Valentina, and Zhuge, Ying, "Glioma Grading Clinical and Mutation Features." UCI Machine Learning Repository, 2022. DOI: https://doi.org/10.24432/C5R62J.

[15] Rubini, L, Soundarapandian, P, and Eswaran, P, "Chronic Kidney Disease." UCI Machine Learning Repository, 2015. DOI: https://doi.org/10.24432/C5G020.

[16] L. Tan, Y. Yang, M. Hu, Y. Zhou, and D. Wu, "Fraim: A feature importance-aware incentive mechanism for vertical federated learning," in *International Conference on Algorithms and Architectures for Parallel Processing*, pp. 132–150, Springer, 2023.

[17] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural computing and applications*, vol. 24, p. 175–186, 2014.

[18] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[19] A. d'Aspremont, F. Bach, and L. El Ghaoui, "Optimal solutions for sparse principal component analysis.," *Journal of Machine Learning Research*, vol. 9, no. 7, 2008.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.