Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Mining Distributed Data using Vertical Federated Learning Review

Manaaf Abdulredha Yassen, DR. Lamia AbedNoor Muhammed*

College of Computer Sciences and Information Technology, University of Al-Qadisiyah, Al-Diwaniyah, Iraq

Email: lamia.abed@qu.edu.iq. manaf.yassen@qu.edu.iq

ARTICLE INFO

Article history:
Received: 02 /09/2022
Rrevised form: 05/10/2022
Accepted : 09 /10/2022
Available online: 01/12/2022

Keywords:

federated learning

vertical federated learning

Data mining

Distributted Data

ABSTRACT

Federated Learning was designed to allow collaborative learning without revealing raw data as worries about machine learning privacy grew. Vertical Federated Learning (VFL) may be utilized for a distributed dataset with the same sample ID space but differs in feature space. And may be used in a wide variety of real-world contexts when parties have the same set of samples but only have partial attributes. Achieving privacy will be a result of this technique's capacity.

Federated Learning enables different repositories of data to learn a shared model collaboratively and at the same time keep the privacy of each one becauseof the increasing awareness of large firms compromising on data security and user privacy. To accomplish federated learning, three learning ways were suggested; horizontal federated learning, vertical federated learning, and transfer federated learning.

Vertical federated learning was adopted when data were spread among different parties. However, each one has different features from the others for identical objects. This paper is related to this type of federated learning.

To maximize model performance while maintaining the privacy of dispersed data, we'll create a framework based on vertical federated learning and suitable techniques.

https://doi.org/ 10.29304/jqcm.2022.14.4.1081

1. Introduction

Companies and some government organizations that collect personal information are subject to new limits and compliance obligations due to recent data privacy laws and regulations, including the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act. As more and more policy and commercial choices are being made using machine learning models, there is a growing need for larger and more thorough datasets. Because of privacy concerns, many applications must deal with data about entities scattered among several data parties. With a computing idea called federated learning (FL), various data parties may pool

∗Corresponding author

Email addresses:

Communicated by ‘sub etitor’

their data to train actionable models in a collaborative setting without disclosing any of their data. Yet, attackers can still deduce whether or not a user's data is in the training set, meaning that FL alone cannot give any verifiable privacy guarantee.[1]

Horizontal federated learning (HFL) is the most typical environment in which FL is implemented. It presupposes that all local datasets have the same attributes but that the data parties contain data from different users. That's analogous to slicing a master data set in half horizontally and handing each half to a separate group. In preparation for the next training iteration, the global model is updated on a centralized server and disseminated to all users. In this study, we examine the vertical federated learning (VFL) context, which is also quite significant[2].

In contrast to the HFL, all participants share data from the same pool of users, but their data attributes are distinct. Multiple users can train a standard model using federated learning without collecting their information. For data privacy, aggregated locally computed updates are used to train models, and no client data is transported elsewhere. Two data sets have the same sampling ID space but differ in feature space in a vertical scenario. [3]

2 Methodology

2.1 Federated Learning

Google was the first to introduce the idea of federated learning. It is a machine learning algorithm approach that covers the following features: Two or more participants work together to train a global model; each node has some local data which can be used to train the global model; data is stored locally by participants all through global model training; privacy protection can be used to avoid privacy leakage throughout local parameter communicating; the accuracy of the federated model is an optimal approximation of the accuracy of the ideal model built from centralized data. The utility of the global model and privacy concerns in the communication process are two implied considerations [4].

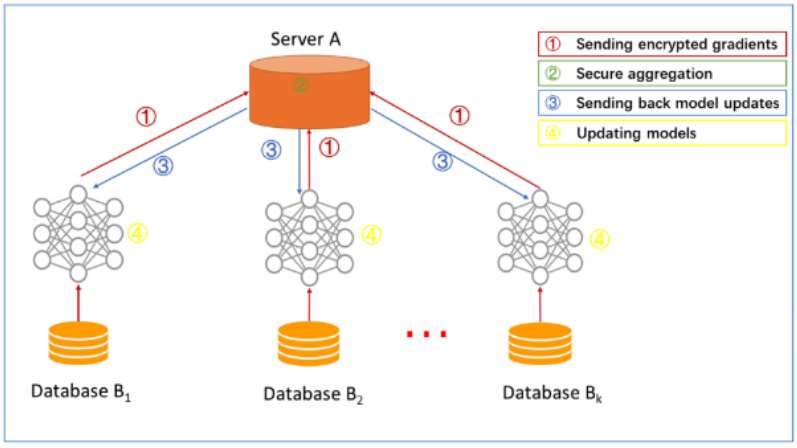
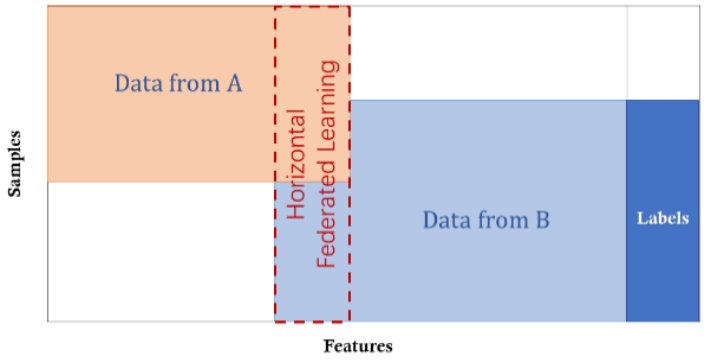


Fig (1) Federated Learning Concept[5]

2.2 Categorizations of federated learning

2.2.1 Horizontal Federated Learning

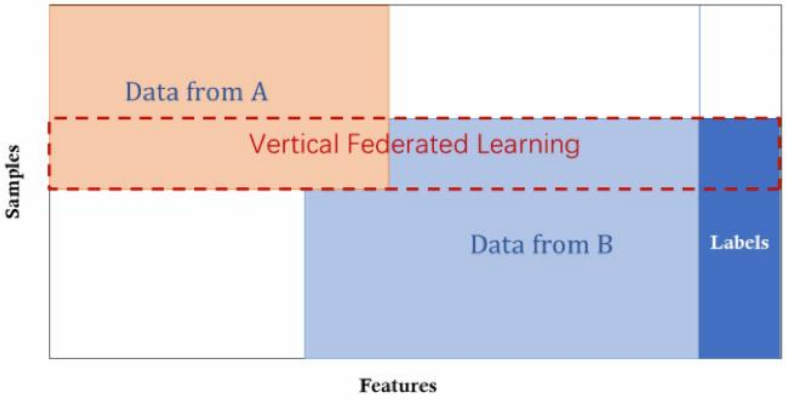
When the user characteristics of the two datasets conflict considerably, but the users do not, horizontal federated learning (HFL) is the best option. It is possible to divide datasets horizontally (by user dimension) and then extract data from which user characteristics are similar but not identical for training. Alternatively, the data in different rows all share the same data properties (aligned by user features). Consequently, more people may be included in the sample using horizontal federated learning. For example, there are two unique suppliers of the same service in different locations, each with a distinct user base. As a result, the user characteristics of both firms are identical [6].



Fig(2) Horizontal Federated Learning[5]

2.2.2 Vertical Federated Learning

Vertical federated learning (VFL) or feature-based learning is applicable when two data sets share the same sample ID space but differ in feature space [5]. Several parties have access to the same data collection, but each party only has access to a distinct subset of features. Many industries, like healthcare, finance, and banking, value privacy above all else, which is why vertical federated learning is gaining more and more traction. As a result, we will concentrate on vertical federated learning in this work.[3]



FIG(3) VERTICAL FEDERATED LEARNING[5]

2.2.3 Federated Transfer Learning

In contrast to horizontal and vertical federations, federated transfer learning is a unique form of federated learning. The feature space of the two datasets differs in federated transfer learning. Accordingly, this holds for data acquired from various organizations with similar traits. There is just a small amount of feature overlap between them due to the differences in the businesses they serve. If you have a global firm, this is true. The outcome is a wide range of datasets with varying sample sizes and feature spaces. For transfer learning to be successful, knowledge from other sections of the target domain must be used. Transfer learning teaches a dummy how to use one product and then apply what they learned to another [7].

2.3 Definition of Vertical Federated Learning

VFL setting. There are two main types of FL configurations: horizontal and vertical. In this section, we explain the fundamentals of VFL, and For the sake of brevity, we'll refer to a whole dataset as D = (I, X, Y), where I, X, and Y stand for the sample ID space, the feature space, and the label space, respectively. That there are two datasets used for vertical federated learning, $D_1 = (I_1, X_1, Y_1)$, $D_2 = (I_2, X_2, Y_2)$, satisfying[8]

$$X_1 \neq X_2 , Y_1 = Y_2 , I_1 = I_2 .$$

3- Modren studies in Vertical Federated Learning :

Jiankai Sun and his colleagues offer "a vertically federated learning" scope that is instituted "Private Set Union (PSU)," which allows every client to keep critical belonging data private. Propose methods for generating artificial attributes and labels for instances that contribute to the union but not the junction. Criteo and Avazu are two real-world datasets that can be used. Criteo advertising with roughly 45 million user clicks records. Avazu has around 40 million entries. Experimental results show that our technique can protect intersecting membership without significantly losing model performance [9].

Depending on the LSTM fault classification network, **Zhili Ma and his colleagues** present a vertical federated learning system (LstFcFedLear). Used the Fault Type of firefighting (FTFF) dataset from China State Grid Gansu Electric Power Company's Firefighting Internet of Things platforms database to solve the problem of fault type classification for fire facilities. The experimental results suggest that the LstFcFedLear model is an excellent way to anticipate faults and that the results are equivalent to the baseline. Compared to the other three approaches, it is more accurate. LstFcFedLear has an accuracy of 94.6 percent, which is 8.03 percent greater than the average of KNN, SVM, and CNN. [10].

Nick Angelou and his research team built an open-source library for PSI and PSI-Cardinality protocols. Combining standard DDH-based PSI protocols with Bloom filter compression to reduce connections in the non-symmetric scenario has been developed..... According to the results, this library is very competitive in terms of runtime and communication. Simultaneously, it is adaptable enough to work with various systems and languages, including browsers. Use our library in the following two scenarios: tracking contacts while protecting the privacy and (ii) machine learning on vertically partitioned data that is privacy-preserving but compatible with present methods [11].

Reverse sum attack and reverse multiplication assault are two practical attacks devised by **Haiqin Weng and his research team**, neither of which will alter the learned model's correctness. The adversary's hostile attitude does not deviate from the protocol specification and crumbles any accuracy of the target model.

The attack is simple - the adversary needs little prior knowledge about the data diversions protocol. This shows that attacks are efficient and evasive. We also demonstrate that the stolen data is just as effective as the raw training data to train a different classifier—the Author Talks about potential countermeasures and their difficulties [12].

Researchers and Xiao Han presented FedValue, a privacy-preserving data valuation method for vertical FL problems without models. FedValue's approach to game theory allows it to examine the data values of many different parties using a unique information-theoretic measure called Shapley CMI. The Shapley-CMI calculation is made possible by a server-assisted federated computation technique, which prevents data leaking for all participants. In addition, we'll discuss numerous ways to speed up Shapley-CMI calculation in practice. It's critical, yet evaluating the parties' data is a challenging FL problem. Both running-specific models and task-agnostic data valuation methods are used in the literature. Trials on six publicly available datasets demonstrate the efficacy and efficiency of FedValue in valuing vertical FL employment data. A model-free metric like Shapley-CMI performs as well as a model-based metric like an ensemble of good-performing models [13].

Christopher Briggs and his colleagues proposed an asymmetrical vertical federated learning (AVL) model (FL+HC). They explain how item IDs may be secured. PSI protocol is modified for an asymmetrical ID alignment phase in an asymmetrical vertical federated learning system. The new protocol now has a Pohlig-Hellman realization. There is also a real-with-dummy approach to federated asymmetrical model training. It's shown in the form of a federated logistic regression algorithm. This was conducted using the MNIST2 dataset, with 60000 samples and 784 attributes. [8].

Wenjie Song and his research group present a new vertical federated learning framework based on the DFP and the BFGS (abbreviated as BDFL), which apply to logistic regression. To tackle the problem of non-iid data and inefficient communication between the client and the server, the Experiment with real datasets to see how effective the BDFL framework is. PDF has also demonstrated that it can meet the following two multi-party modelling premises:1) Ensure data privacy is not compromised; 2) One has merely data and no labelling. The other has data as well as a label. In addition, the model's convergence speed and accuracy are superior to existing methods [14].

DVFL, a new vertical federation learning approach proposed by **Yuzhi Liang and Yixiang Chen**, adapts to dynamic data distribution changes through knowledge distillation. To improve data security and model performance, most computations in DVFL are performed locally. DVFL was formerly utilized to solve the problem. Modern VFL approaches are primarily employed in static circumstances where both the active and passive parties have all of the data from the start and will not change. However, data in real life is frequently dynamic. Breast Cancer Wisconsin (BCW), Default of Credit Card Clients (DCC), Epsilon (ESP), and Human Activity Recognition (HAR) were the four datasets employed in his research. Since BCW and DCC are labeled unbalanced datasets, the results demonstrate that Marco-P, Marco-R, and Marco-F1 are employed as evaluation metrics. In these studies, 5-fold cross-validation is also used [15].

According to **Xiang Ni and his team of researchers**, FedVGCN is a model of federated GCN learning that can be applied to current GCN models while still protecting node privacy in a vertically partitioned environment. Describe a novel approach to securing privacy while maintaining accuracy by employing additively homomorphic encryption (HE). Datasets. Three primary benchmark datasets are used to evaluate the performance of GNN: Cora, PubMed, and Citeseer. The method dramatically outperforms GNN models trained on isolation data for three benchmark datasets and is equivalent to standard GNN models trained on mixed plaintext data. [3].

Dataset	Cora	Pubmed	Citeseer
$GraphSage_A$	0.5222	0.6936	0.4630
$GraphSage_B$	0.4867	0.6801	0.5510
FedVGraphSage	0.6770	0.7830	0.6820
$GraphSage_{A+B}$	0.7080	0.7890	0.6983

There has been increased interest in this defense concept, and researchers have proposed an enhanced data leaking attack with a theoretical justification for obtaining batch information via shared aggregating gradients. The proposed method is "catastrophic information leaking in vertical federated learning" ("CAFE"). Our results suggest that ("CAFE") is more successful for large-scale leakage attacks when using vertical FL settings than other data leakage attacks. A credible ("CAFE") mitigation plan should also be provided. Due to our investigation, researchers have uncovered new and real data leakage threats in some cases.. [16].

For multi-class VFL situations involving several parties, **Siwei Feng and Han Yu** suggest the MMVFL framework (multi-participant Multi-class Vertical Federated Learning). MMVFL incorporates a feature selection method to compare its performance to supervised feature selection and MVL-based approaches to explain its effectiveness. According to experiments on real-world datasets, using two benchmark MVL datasets: Handwritten and Caltech7, the former of which has five views and the latter of which contains 6, corresponding to 5 and 6 VFL participants [17].

Wenjing Fang and his colleagues hope to create a large-scale secure XGB (SS-XGB) in a vertically federated learning environment. We protect your personal information in three ways. Specifically, we use safe multi-party computation approaches to avoid leaking intermediate information during training. We store the output model in a distributed manner to limit information leakage. We present a unique algorithm for safe XGB prediction with the distributed model. Furthermore, by introducing secure permutation algorithms, the system can scale to massive datasets and enhance training efficiency. Comprehensive experiments are carried out on both public and real-world datasets, and the results indicate that our suggested XGB models provide cutthroat accuracy and practical performance [18].

Wensheng Xia and his colleagues offer Cascade Vertical Federated Learning (CVFL), a new vertical federated learning architecture that utilizes all horizontally partitioned labels while maintaining privacy. To address the straggler problem, devise a unique optimization target that might increase the contribution of stragglers to trained models. Most available vertical federated learning algorithms still face two significant problems in real-world applications. First, many current vertical federated learning methods assume that at least one party has the whole set of labels for all data samples, which isn't always true in practice, especially when labels are horizontally partitioned, and the parties have only partial labels. Current vertical federated learning approaches can only use partial labels, potentially resulting in an insufficient model update in end-to-end backpropagation. Second, each party's computational and communication resources differ. To rigorously verify the efficiency of CVFL, perform a series of qualitative trials. By centralizing training, CVFL can attain equivalent performance (for example, accuracy for classification tasks). Compared to merely applying the asynchronous aggregation approach during training, the new optimization target can further mitigate the straggler problems [3].

Shengwen Yang and his colleagues provide a technique for vertical federated learning using parallel distributed logistic regression. The system relies on the parameter server architecture and tries to increase model training by employing a cluster of servers in cases with a huge amount of training data. The system's efficiency and scalability were further tested on two datasets (sparse and dense), with experimental findings demonstrating the system's efficiency and scalability. The experimental results demonstrate that: 1) the learned models perform well as measured by AUC; 2) the model speedily converges to a steady-state after a few iterations. And 3) the system is flexible and has sublinear speedup. [19].

Qingsong Zhang and his colleagues offer an AsySQN (asynchronous stochastic quasi-Newton) architecture for VFL, which includes three algorithms: AsySQN-SGD, -SVRG, and -SAGA. In reality, the suggested AsySQN-type algorithms that scale descent steps by approximation, the AsySQN architecture uses estimated second-order information to drastically decrease the number of communications rounds, resulting in

lower communication costs. used eight datasets, (UCICreditCard) and (GiveMeSomeCredit) are from Kaggle1, while (news20), (w8a), (rcv1), (a9a), (epsilon), and (mnist) are from LIBSVM2 [20].

Changxin Liu and his colleagues present a fair VFL system. each data party makes several parallelized local modifications per connection round to effectively decrease the number of communications rounds, to solve it in a Vertical federated approach. Rigorous tests on three benchmark datasets show that our strategy outperforms others regarding training fair models. ADULT is a dataset for adults. In the experiment, 40,000 out of 45,222 data points are uniformly sampled for training, and the rest, 5,222 data instances, are used for testing. The COMPAS dataset contains 5,278 data instances, of which 4,800 are uniformly sampled as training data, and the remainder 478 are used as testing data in the trials. There are 1994 data cases in the Community and Crime (C&C) dataset; uniformly sample 1,200 data samples as training data and use the remainder of 794 as testing data [21].

Hangyu Zhu and his colleagues offer PIVODL, a safe vertical FL framework for training GBDT with data labels spread across several devices. The experimental results reveal that the suggested PIVODL has little information leaking and model performance impairment. PIVODL's connection cost and label predictive inference. The dataset used 1- Credit card: It comprises 30,000 raw data with 23 attributes. 2- Bank marketing: There are 45211 instances and 17 characteristics. 3- Appliances energy prediction: It's a regression dataset with 19735 data points and 29 attributes and Analyzes the changes in learning performance as a function of the number of participating clients.[22].

Zhaomin Wu and his colleagues developed FedSim. FedSim improves VFL performance by directly employing the similarities calculated in PPRL and avoiding the classification process. It constructs algorithms presuming that data from various parties have been linked. Make three observations based on the results. FedSim outperforms or outperforms all baselines and significantly outperforms Combined in synthetic datasets with small-scale noise sources [23].

A new strategy for training VFL models with numerous data and label owners has been proposed by **Vaikkunth Mugunthan and his researcher**, Multi Vertical Federated Learning (Multi-VFL). This proposed structure allows for the training and discovery of optimum models by different entities without the requirement to exchange data. Split learning and adaptive federated optimizers are two of the framework's tools for dealing with this problem. Adaptive optimizers outperformed and converged quicker than the FedAvg method on the MNIST and FashionMNIST datasets for different non-IID label distribution circumstances, with adaptive optimizers improving accuracy by 2-3 percent in the Fashion MNIST dataset. The FashionMNIST dataset also reveals that FedAvg fails to converge [2].

Daniele Romanini and his team provide PyVertical, an innovative paradigm for vertical federated learning employing partitioning neural networks. You may use Private Set Intersection on IDs linked to data points to link items shared across different datasets' divisions. PyVertical, as a consequence, may be used to study neural-network-based data samples from two data consumers and a data scientist. VFL. Using PSI to resolve data themes across datasets is acceptable and effective; a dual-headed model trained on a vertically partitioned MNIST dataset was excellent. [24].

FedV, an approach for secure gradient computing in vertical settings of numerous commonly used ML models such as linear models, logistic regression, and support vector machines, is proposed by **Runhua and his research group**. Existing vertical FL techniques necessitate repeated peer-to-peer connections between participants, resulting in extensive training durations, and are limited to (roughly) linear models and only two parties. Experimentally illustrate the applicability for various ML models and exhibit a decrease in training time

of 10%-70 percent and data transfer of 80%-90 percent compared to state-of-the-art techniques. The result is two FedV models: 1) a logistic regression model trained, 2) a logistic regression model of Taylor series estimation, which decreases the logistic regression model to a linear model. Including logistic regression and

SVMs, without the requirement for any approximations. as a result, it can be used in difficult situations when a client cannot maintain communication during the training process [25].

Qingsong Zhang and his colleagues offer a revolutionary VFL framework that includes a new backward updating technique, a bilevel asynchronously parallelism construction (VFB2), and three new algorithms: VFB2 -SGD, -SVRG, and -SAGA. In real-world VFL applications, just one or a few parties often hold labels, making it difficult for all parties to train the model collectively without exposing personal information. Meanwhile, most existing VFL algorithms are stuck in synchronized calculations, resulting in inefficiency in real-world applications. As a result, deduce the theoretical findings of these three algorithms' converging ratios under both substantially convex and nonconvex situations. Further, demonstrate VFB2's protection using semi-honest threat models. Extensive testing on benchmark datasets has shown that our techniques are fast, flexible, and lossless [26].

Authors	year	Proposed technique	Purpose
Jiankai Sun and the others [9]	**2021**	**a VFL framework based on Private Set Union (PSU)**	**propose a novel VFL framework that addresses the intersection membership leakage problem, which currently prevents many privacy-sensitive organizations from adopting VFL**
Zhili Ma and other teams. [10]	**2021**	**vertical federated learning framework based on LSTM fault classification network (LstFcFedLear)**	**framework is that it can encrypt and integrate the data on the entire firefighting IoT platform to form a new dataset**
Nick Angelou and other teams[11]	**2020**	**a versatile open-source library for asymmetric private set intersection (PSI) and PSI-Cardinality (PSI-C)**	**privacy-preserving machine learning on vertically partitioned data and a privacy-preserving contact tracing protocol that is compatible with existing approaches.**
Haiqin Weng and his teams [12]	**2020**	**two simple yet effective attacks, reverse multiplication attack and reverse sum attack**	**adversaries that participate in the training of a distributed ML model, but do not stray from the stated protocol, seek to deduce private training data from the lawfully received information.**
Researchers and Xiao Han[12]	**2021**	**design and implement two practical attacks, reverse sum attacks, and reverse multiplication attacks.**	**Adversaries that participate in a distributed ML model training, but do not stray from the stated protocol, seek to deduce private training data from the lawfully received information.**

References

[1] Z. Li, T. Wang, and N. Li, "Differentially Private Vertical Federated Clustering." 2022, [Online]. Available: http://arxiv.org/abs/2208.01700.

[2] V. Mugunthan, P. Goyal, and L. Kagal, "Multi-VFL: A Vertical Federated Learning System for Multiple Data and Label Owners." 2021, [Online]. Available: http://arxiv.org/abs/2106.05468.

[3] X. Ni, X. Xu, L. Lyu, C. Meng, and W. Wang, "A Vertical Federated Learning Framework for Graph Convolutional Network," Jun. 2021, doi: 10.48550/arXiv.2106.10056.

[4] J. Zhao, K. Mao, C. Huang, and Y. Zeng, "Utility Optimization of Federated Learning with Differential Privacy," *Discrete Dynamics in Nature and Society*, vol. 2021. 2021, doi: 10.1155/2021/3344862.

[5] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2. 2019, doi: 10.1145/3298981.

[6] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Syst.*, vol. 216, no. March, p. 106775, Mar. 2021, doi: 10.1016/j.knosys.2021.106775.

[7] S. Saha and T. Ahmad, "Federated transfer learning: Concept and applications," *Intelligenza Artificiale*, vol. 15, no. 1. pp. 35–44, 2021, doi: 10.3233/IA-200075.

[8] Y. Liu, X. Zhang, and L. Wang, "Asymmetrical Vertical Federated Learning." 2020, [Online]. Available: http://arxiv.org/abs/2004.07427.

[9] J. Sun *et al.*, "Vertical Federated Learning without Revealing Intersection Membership." 2021, [Online]. Available: http://arxiv.org/abs/2106.05508.

[10] X. Zhang, Z. Ma, A. Wang, H. Mi, and J. Hang, "LstFcFedLear: A LSTM-FC with Vertical Federated Learning Network for Fault Prediction," *Wireless Communications and Mobile Computing*, vol. 2021. 2021, doi: 10.1155/2021/2668761.

[11] N. Angelou *et al.*, "Asymmetric Private Set Intersection with Applications to Contact Tracing and Private Vertical Federated Machine Learning." 2020, [Online]. Available: arxiv:2011.09350.

[12] H. Weng, J. Zhang, F. Xue, T. Wei, S. Ji, and Z. Zong, "Privacy Leakage of Real-World Vertical Federated Learning," *arxiv*, 2020, [Online]. Available: https://arxiv.org/pdf/2011.09290.

[13] X. Han, L. Wang, and J. Wu, "Data Valuation for Vertical Federated Learning An Information-Theoretic Approach.pdf." 2021, doi: 2112.08364.

[14] S. WenJie and S. Xuan, "Vertical federated learning based on DFP and BFGS." 2021, [Online]. Available: http://arxiv.org/abs/2101.09428.

[15] Y. Liang and Y. Chen, "DVFL: A Vertical Federated Learning Method for Dynamic Data." 2021, [Online]. Available: http://arxiv.org/abs/2111.03341.

[16] X. Jin, P.-Y. Chen, C.-Y. Hsu, C.-M. Yu, and T. Chen, "CAFE_Catastrophic_Data_Leakage_in_Vertical_Federat (1).pdf." 2021, doi: 2110.15122.

[17] S. Feng and H. Yu, "Multi-Participant Multi-Class Vertical Federated Learning." 2020, [Online]. Available: http://arxiv.org/abs/2001.11154.

[18] W. Fang *et al.*, "Large-scale Secure XGB for Vertical Federated Learning," in *International Conference on Information and Knowledge Management, Proceedings*, Oct. 2021, pp. 443–452, doi:

10.1145/3459637.3482361.

[19] S. Yang, B. Ren, X. Zhou, and L. Liu, “Parallel Distributed Logistic Regression for Vertical Federated Learning without Third-Party Coordinator.” 2019, [Online]. Available: http://arxiv.org/abs/1911.09824.

[20] Q. Zhang *et al.*, “AsySQN: Faster Vertical Federated Learning Algorithms with Better Computation Resource Utilization,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2021, pp. 3917–3927, doi: 10.1145/3447548.3467169.

[21] C. Liu, Z. Zhou, Y. Shi, J. Pei, L. Chu, and Y. Zhang, “Achieving Model Fairness in Vertical Federated Learning.” 2021, [Online]. Available: http://arxiv.org/abs/2109.08344.

[22] H. Zhu, R. Wang, Y. Jin, and K. Liang, “PIVODL: Privacy-preserving vertical federated learning over distributed labels.” 2021, [Online]. Available: http://arxiv.org/abs/2108.11444.

[23] Z. Wu, Q. Li, and B. He, “Exploiting Record Similarity for Practical Vertical Federated Learning.” 2021, [Online]. Available: http://arxiv.org/abs/2106.06312.

[24] D. Romanini *et al.*, “PyVertical: A Vertical Federated Learning Framework for Multi-headed SplitNN.” 2021, [Online]. Available: http://arxiv.org/abs/2104.00489.

[25] R. Xu, N. Baracaldo, Y. Zhou, A. Anwar, J. Joshi, and H. Ludwig, “FedV Privacy-Preserving Federated Learning over Vertically.pdf,” 2021, [Online]. Available: https://arxiv.org/pdf/2103.03918.

[26] Q. Zhang, B. Gu, C. Deng, and H. Huang, “Secure Bilevel Asynchronous Vertical Federated Learning with Backward Updating,” Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.00958.