CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

# Journal Paper

## A Novel Framework of Horizontal-Vertical Hybrid Federated Learning for EdgeIoT

**Kai Li***
**Yilei Liang**
**Xin Yuan**
**Wei Ni**
**Jon Crowcroft**
**Chau Yuen**
**Ozgur B. Akan**

*CISTER Research Centre

CISTER-TR-250202

2025

# A Novel Framework of Horizontal-Vertical Hybrid Federated Learning for EdgeIoT

Kai Li*, Yilei Liang, Xin Yuan, Wei Ni, Jon Crowcroft, Chau Yuen, Ozgur B. Akan

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: kai@isep.ipp.pt, yl841@cam.ac.uk, xin.yuan@data61.csiro.au, Wei.Ni@data61.csiro.au, Jon.Crowcroft@cl.cam.ac.uk, yuenchau@sutd.edu.sg, oba21@cam.ac.uk

https://www.cister-labs.pt

## Abstract

This letter puts forth a new hybrid horizontal-vertical federated learning (HoVeFL) for mobile edge computing-enabled Internet of Things (EdgeIoT). In this framework, certain EdgeIoT devices train local models using the same data samples but analyze disparate data features, while the others focus on the same features using non-independent and identically distributed (non-IID) data samples. Thus, even though the data features are consistent, the data samples vary across devices. The proposed HoVeFL formulates the training of local and global models to minimize the global loss function. Performance evaluations on CIFAR-10 and SVHN datasets reveal that the testing loss of HoVeFL with 12 horizontal FL devices and six vertical FL devices is 5.5% and 25.2% higher, respectively, compared to a setup with six horizontal FL devices and 12 vertical FL devices.

© 2025 CISTER Research Center
www.cister-labs.pt

# A Novel Framework of Horizontal-Vertical Hybrid Federated Learning for EdgeIoT

Kai Li, *Senior Member, IEEE,* Yilei Liang, Xin Yuan, *Senior Member, IEEE,* Wei Ni, *Fellow, IEEE,* Jon Crowcroft, *Fellow, IEEE,* Chau Yuen, *Fellow, IEEE,* and Ozgur B. Akan, *Fellow, IEEE*

***Abstract*—This letter puts forth a new hybrid horizontal-vertical federated learning (HoVeFL) for mobile edge computing-enabled Internet of Things (EdgeIoT). In this framework, certain EdgeIoT devices train local models using the same data samples but analyze disparate data features, while the others focus on the same features using non-independent and identically distributed (non-IID) data samples. Thus, even though the data features are consistent, the data samples vary across devices. The proposed HoVeFL formulates the training of local and global models to minimize the global loss function. Performance evaluations on CIFAR-10 and SVHN datasets reveal that the testing loss of HoVeFL with 12 horizontal FL devices and six vertical FL devices is 5.5% and 25.2% higher, respectively, compared to a setup with six horizontal FL devices and 12 vertical FL devices.**

***Index Terms*—Internet of Things, Edge Computing, Hybrid Federated Learning, Horizontal and Vertical, Non-IID Data**

## I. Introduction

Federated learning (FL) presents a transformative solution to mobile edge computing-enabled Internet of Things (EdgeIoT). In an EdgeIoT-empowered hospital, each institution holds vast amounts of sensitive patient data collected from IoT devices, which can enhance the predictive capabilities of medical artificial intelligence (AI) models - critical for diagnoses, treatment planning, and outcomes prediction [1]. However, sharing patient data across institutions can pose privacy risks and violate regulations [2], such as The Health Insurance Portability and Accountability Act of 1996 in the United States or the Data Protection Act 2018 in the United Kingdom.

EdgeIoT devices might have distinct sets of clients but collect unified data features on those patients, such as demographic information, medical histories, and treatment outcomes. Horizontal FL (HFL) can enable clinics to collaboratively train a machine learning model, which captures the same features of non-independent and identically distributed (non-IID) data [3]. FL can also perform vertically, namely, vertical FL (VFL), where EdgeIoT trains on disparate types of information (i.e., data features) from the same clients [4].

In this letter, we propose a new hybrid horizontal-vertical FL (HoVeFL) for EdgeIoT, which enables certain devices to focus on the same features across non-IID samples, while others focus on different features of the same samples. For instance, some EdgeIoT devices may analyze particular attributes of a set of data, whereas other devices examine entirely different attributes pertaining to those same data samples. Although the data samples might involve the same participants or entities, the features under analysis vary across devices. Moreover, HoVeFL allows certain devices to concentrate on identical types of features but under non-IID conditions. For example, some devices might focus on a specific category of attributes; however, the data they possess may come from distinct subsets with diverse characteristics. Even though the targeted features are the same, the actual distributions and contexts of the data samples can differ markedly among the EdgeIoT devices.

HoVeFL has potential applications in settings where data is diverse and requires privacy protections, such as smart health systems with inherently heterogeneous devices and data sources. EdgeIoT devices can vary widely in functionality, for instance, some devices may capture certain types of data under different conditions, while others handle more specialized data formats. In HoVeFL, specific devices can participate in HFL, focusing on shared datasets collected under varying conditions, thereby addressing the non-IID nature of the data. Other devices can participate in VFL, analyzing different datasets drawn from the same subjects. By employing HoVeFL, smart health systems can facilitate comprehensive models that leverage diverse data forms without compromising patient privacy.

## II. Related Work

A hybrid FL framework was examined in [5] to predict the load of transformer stations in an urban setting, where the training datasets consist exclusively of load consumption and related features. Given the homogeneity and IID nature of the data, vertically partitioned datasets from transformer stations across different districts can be synchronized prior to training by aligning sample timestamps. In [6], hybrid FL was characterized by each user's dataset containing partial features, with users holding IID data. In this framework, a

K. Li is with the Internet of Everything (IoE) Group, Department of Engineering, University of Cambridge, CB3 0FA Cambridge, UK, and also with Real-Time and Embedded Computing Systems Research Centre (CISTER), Porto 4249–015, Portugal (E-mail: kaili@ieee.org).

Y. Liang and J. Crowcroft are the Department of Computer Science and Technology, University of Cambridge, CB3 0FA Cambridge, UK (E-mail: yl841@cam.ac.uk & Jon.Crowcroft@cl.cam.ac.uk).

X. Yuan and W. Ni are with Commonwealth Scientific and Industrial Research Organization (CSIRO), Sydney, NSW 2122, Australia, and the School of Computing Science and Engineering, University of New South Wales (E-mail: {xin.yuan, wei.ni}@data61.csiro.au).

C. Yuen is with Nanyang Technological University, SG 639798, Singapore (E-mail: chau.yuen@ntu.edu.sg).

O. B. Akan is with the Internet of Everything (IoE) Group, Division of Electrical Engineering, Department of Engineering, University of Cambridge, CB3 0FA Cambridge, UK, and also with the Center for neXt-generation Communications (CXC), Koç University, 34450 Istanbul, Turkey (E-mail: oba21@cam.ac.uk).

The source code of this work is released on GitHub: https://github.com/YukariSonz/VHFL_SplitNN.

knowledge transfer algorithm was developed, enabling users to utilize their partial features or request additional features from the server to enhance their local models. A concurrent horizontal and vertical FL scheme was introduced in [7] to tackle non-IID features in federated settings. To aggregate the various models generated by different users with non-IID data, the study utilized fuzzy cognitive maps based on fuzzy cause-effect relationships.

The authors of [8] introduced a split learning method to improve data privacy and model robustness. The split learning combines HFL, which is parallel processing among distributed devices, and the split learning, which is network splitting into client-side and server-side sub-networks during training. In [9], a split FL algorithm was developed to reap parallel model training and the model splitting structure of split learning. A multi-arm Bandit algorithm was studied to select user devices and the size of local model updates, balancing the channel qualities and the importance of local model updates.

The primary research gap addressed by the proposed HoVeFL framework lies in its integration of horizontal and vertical strategies within a single FL system. Existing models typically focus on either HFL, which processes shared data types across different populations, or VFL, which analyzes different data types for the same population but cannot leverage their respective strengths. HoVeFL bridges this gap by enabling simultaneous horizontal and vertical learning on heterogeneous data types and varied populations. This hybrid approach allows for more comprehensive model training that can adapt to non-IID data across devices, addressing the challenges of data diversity and distribution discrepancies more effectively than single HFL or VFL strategies.

Unlike the approaches in [5] and [6], our proposed HoVeFL is a new hybrid FL framework. HoVeFL enables some devices to concentrate on identical features across datasets with non-IID samples, while others focus on different features within datasets that are IID. This unique configuration allows for a more flexible and efficient learning process by accommodating both horizontal and vertical data partitioning simultaneously. Although the study in [7] incorporates non-IID data, the model aggregation assigns uniform weights to all local models, without differentiating those originating from HFL or VFL. In contrast, our proposed HoVeFL focuses on a different FL scenario that allows the server to selectively adjust weights for the aggregated local models. The server can assign the local models with higher relevance a greater weight during model aggregation. In addition, not only does our study present the conceptual framework of HoVeFL but also rigorously investigates its convergence properties under non-IID data conditions.

## III. Formulation of HoVeFL

This section presents the training of the local and global models of HoVeFL in EdgeIoT for image classification as an example. Fig. 1 presents a HoVeFL training process with $N$ user devices, where data feature $i$ ($i \in [1, I]$) is trained by $N_i$ devices and data samples $j$ are trained by $N_j$ devices. Each device $n \in [1, N]$ has $n_{i,j}(t)$ with $j$ data samples and $i$ features at the $t$-th round of HoVeFL. Let $\boldsymbol{m}_{i,j}^n(t) \in \mathbb{R}^{1\times M}$

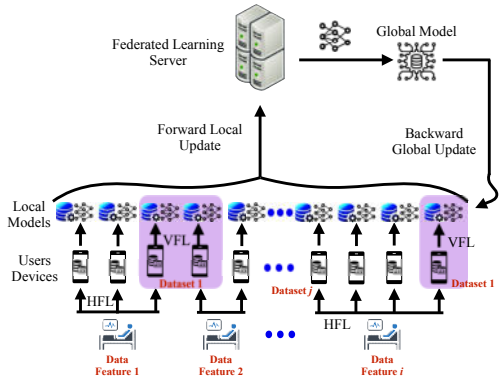


Fig. 1. A training process of the local and global models in HoVeFL, where each data feature $i$ ($i \in [1, I]$) is trained by $N_i$ devices. Moreover, each device has data samples $j$, thus, we have $n_{i,j} \in \mathbf{N_i} \cap \mathbf{N_j}$.

denote the local model update in the $t$-th round. The training loss function of device $n$ is denoted by $f(\boldsymbol{m}_{i,j}^n(t))$ and measures approximation errors based on training datasets in the $t$-th round. For example, the loss function can be modeled as linear or logistic regressions [10].

Let $D_j^n(t)$ denote the data samples $j$ at device $n$. The local loss function of HoVeFL at device $i$ for the $t$-th round is

$$F(\boldsymbol{m}_{i,j}^n(t)) = \frac{1}{D_j^n(t)} \sum_{i=1}^{D_j^n(t)} f\left(\boldsymbol{m}_{i,j}^n(t)\right) + \alpha\zeta\left(\boldsymbol{m}_{i,j}^n(t)\right), \quad (1)$$

where $\zeta(\cdot)$ is a regularizer function mitigating the effect of local training noise and potential overfitting due to overlapping input features among different devices. $\zeta(\cdot)$ contributes to the generalization ability of the model across the network. $\alpha \in [0, 1]$ is a given coefficient.

Minimizing the local loss function in (1) leads to an optimal local model $\boldsymbol{m}_{i,j}^{n*}(t)$ at device $n$, where

$$\boldsymbol{m}_{i,j}^{n*}(t) = \arg \min_{\forall \boldsymbol{m}_{i,j}^n(t)} F(\boldsymbol{m}_{i,j}^n(t)). \quad (2)$$

With the learning rate $\mu_t$, the local model of device $i$ is updated using Stochastic Gradient Descent (SGD) for $T_L$ local iterations, throughout the $t$-th round:

$$\boldsymbol{m}_{i,j}^n(t) \leftarrow \boldsymbol{m}_{i,j}^n(t) - \mu_t \nabla F(\boldsymbol{m}_{i,j}^n(t)). \quad (3)$$

After $T_L$ local iterations, the server aggregates $\boldsymbol{m}_{i,j}^n(t), \forall i$ to obtain a global model, denoted by $\boldsymbol{m}_G(t)$, for the $t$-th round. Then, $\boldsymbol{m}_G(t)$ is broadcast to all devices for their training of $\boldsymbol{m}_{i,j}^n(t+1)$, $\forall i$ in the $(t+1)$-th round. The overlapping input features between the devices are handled differently. Although overlapping input features across devices are allowed, each device's participation in either HFL or VFL dictates its method of data integration. For VFL, where the devices might share partially overlapping input features on the same datasets, these features are aligned and integrated vertically to enrich the model's understanding of a device. This is achieved through techniques, such as feature fusion or aggregation, ensuring that all relevant information from overlapping features is effectively utilized. For HFL, the emphasis is on leveraging the same set of features across different devices, and any overlapping features among the devices participating in HFL

$$F(\boldsymbol{m}_G(t)) = \frac{1}{\sum_{i=1}^{I} N_i} \sum_{i=1}^{I} w_i^{n_i} (\boldsymbol{m}_{i,1}^{1_i}(t) + \boldsymbol{m}_{i,2}^{2_i}(t) + \cdots + \boldsymbol{m}_{i,j}^{N_i}(t)) + \frac{1}{\sum_{j=1}^{J} N_j} \sum_{j=1}^{J} (w_1 \boldsymbol{m}_{1,j}^{1_j}(t) + w_2 \boldsymbol{m}_{2,j}^{2_j}(t) + \cdots + w_i^{n_j} \boldsymbol{m}_{i,j}^{N_j}(t)). \tag{4}$$

help enhance the generalizability and robustness of the model across diverse datasets. Thus, the overlap of input features is managed within the distinct protocols of either HFL or VFL, depending on the device's designated learning strategy.

In HoVeFL, $N_i$ users train the local models for the same data feature $i$ but on different data samples, namely, $\boldsymbol{m}_{i,1}^{1}(t), \boldsymbol{m}_{i,2}^{2}(t), \cdots, \boldsymbol{m}_{i,j}^{N_i}(t)$, $i \in [1, I]$. Meanwhile, the other $N_j$ users train the local models based on similar sample $j$ but for different features, i.e., $\boldsymbol{m}_{1,j}^{1}(t), \boldsymbol{m}_{2,j}^{2}(t), \cdots, \boldsymbol{m}_{i,j}^{N_j}(t)$, $j \in [1, J]$. Take a typical aggregation scheme, such as federated averaging (FedAvg) [11], as an example. At the server, a weighted model aggregation aggregates hybrid local models so that the global loss function $F(\boldsymbol{m}_G(t))$ in (4) is minimized, where $w_i^n$ denotes the weight for each feature in the user's dataset as the server may apply a feature selection scheme. Although individual devices may not have a complete set of features to train independent models, they can contribute to a global model that enables independent predictions [12].

While the server handles a more sophisticated aggregation process to integrate horizontal and vertical learning updates, the EdgeIoT devices themselves are not subjected to additional coordination requirements or increased communication overhead. The devices continue to operate as in traditional FL setups, i.e., performing local computations and sending model updates to the server, without needing to manage the complexities introduced by the HoVeFL framework. The computational overhead increases due to the enhanced aggregation process at the server. Since the server is typically equipped with advanced computational capabilities and designed to handle intensive processing tasks, this additional overhead is manageable and does not impede the system's scalability. The practical applicability of HoVeFL is not affected, and the overall efficiency of HoVeFL across EdgeIoT systems remain uncompromised.

## IV. Convergence of HoVeFL

In this section, we investigate the convergence of HoVeFL given non-IID data at the devices.

**Assumption 1.** *$\forall k \in \mathcal{K}$, we make the following assumptions:*

1) *$F(\boldsymbol{m}_G(t))$ fulfills the Polyak-Lojasiewicz requirement [13] with a positive parameter $\rho$, indicating that $F(\boldsymbol{m}_G(t)) - F(\boldsymbol{m}_G^*) \leq \frac{1}{2\rho} \|\nabla F(\boldsymbol{m}_G(t))\|^2$ and $\boldsymbol{m}_G{}^*$ minimizes $F(\boldsymbol{m}_G(t))$, where $\|\cdot\|$ stands for norm operation;*
2) *$F(\boldsymbol{m}_G(0)) - F(\boldsymbol{m}_G(t)^*) = \Theta$, where $\Theta$ is a constant.*

Based on (1) and (4), the global loss function minimized in HoVeFL can be defined as $\min_{\boldsymbol{m}_G(t)} F(\boldsymbol{m}_G(t))$. The gradient descent update is given by

$$F(\boldsymbol{m}_G(t+1)) = F(\boldsymbol{m}_G(t)) - \mu_t \sum_{n=1}^{N} \nabla F(\boldsymbol{m}_G(t)). \tag{5}$$

Assume that the gradients of the loss functions $\nabla F(\boldsymbol{m}_G(t))$ are $L$-Lipschitz continuous [14]. We have

$$\|\nabla F(\boldsymbol{m}_G(t+1)) - \nabla F(\boldsymbol{m}_G(t))\| \leq L \|\boldsymbol{m}_G(t+1) - \boldsymbol{m}_G(t)\|. \tag{6}$$

Given the gradient descent update in (5) and the Lipschitz gradient condition in (6), an inequality that bounds the per-iteration change in the loss function is given by

$$F(\boldsymbol{m}_G(t+1)) \leq F(\boldsymbol{m}_G(t)) - \mu_t \Big( \nabla F(\boldsymbol{m}_G(t)), \sum_{n=1}^{N} \nabla F(\boldsymbol{m}_{i,j}^n(t)) \Big) + \frac{L\mu_t^2}{2} \left\| \sum_{n=1}^{N} \nabla F(\boldsymbol{m}_{i,j}^n(t)) \right\|^2. \tag{7}$$

Furthermore, considering Cauchy-Schwarz and Jensen's inequality (which is used for handling expectations over convex functions) in the gradient descent update can smooth the model updates despite the heterogeneity in $D_j^n(t)$. In particular, the expectation of the convex combination of local losses (which are possibly based on different data distributions) is as good as the convex combination of their expectations, ensuring an averaging effect that is central to the convergence of HoVeFL.

To derive explicit upper bounds of HoVeFL, we apply the Cauchy-Schwarz and Jensen's inequalities to (7), i.e.,

$$\Big( \nabla F(\boldsymbol{m}_G(t)), \sum_{n=1}^{N} \nabla F(\boldsymbol{m}_{i,j}^n(t)) \Big) = \|\nabla F(\boldsymbol{m}_G(t))\|^2, \tag{8}$$

$$\left\| \sum_{n=1}^{N} \nabla F(\boldsymbol{m}_{i,j}^n(t)) \right\|^2 \leq \Big( \|\nabla F(\boldsymbol{m}_G(t))\| + \sigma \Big)^2. \tag{9}$$

Let $\sigma$ denote a variance of the gradients due to non-IID data

$$\sum_{n=1}^{N} \|\nabla F(\boldsymbol{m}_{i,j}^n(t)) - \nabla \overline{F}(\boldsymbol{m}_{i,j}^n(t))\|^2 \leq \sigma^2, \tag{10}$$

where $\overline{F}(\boldsymbol{m}_{i,j}^n(t))$ is the average of the model gradients.

By substituting (8) and (9) into (7), we have

$$F(\boldsymbol{m}_G(t+1)) \leq F(\boldsymbol{m}_G(t)) - \mu_t \|\nabla F(\boldsymbol{m}_G(t))\|^2 + \frac{L\mu_t^2}{2} (\|\nabla F(\boldsymbol{m}_G(t))\| + \sigma)^2. \tag{11}$$

According to Cauchy-Schwarz Inequality, we have

$$F(\boldsymbol{m}_G(t+1)) \leq F(\boldsymbol{m}_G(t)) + (L\mu_t^2 - \mu_t) \|\nabla F(\boldsymbol{m}_G(t))\|^2 + L\mu_t^2 \sigma^2. \tag{12}$$

By subtracting $F(\boldsymbol{m}_G^*)$ from both sides of (12), it follows

$$F(\boldsymbol{m}_G(t+1)) - F(\boldsymbol{m}_G^*) \leq F(\boldsymbol{m}_G(t)) - F(\boldsymbol{m}_G^*) + L\mu_t^2 \sigma^2 + (L\mu_t^2 - \mu_t) \|\nabla F(\boldsymbol{m}_G(t))\|^2. \tag{13}$$

Considering Polyak-Lojasiewicz condition in Assumption 1-1), we can obtain

$$\|\nabla F(\boldsymbol{m}_G(t+1))\|^2 \leq [2\rho(L\mu_t^2-\mu_t)+1]\,\|\nabla F(\boldsymbol{m}_G(t))\|^2 + 2\rho L\mu_t^2\sigma^2. \quad (14)$$

Based on the recurrence expression (14), we can obtain the upper bound of $\|\nabla F(\boldsymbol{m}_G(t))\|^2$, as given by

$$\|\nabla F(\boldsymbol{m}_G(t))\|^2 \leq [2\rho(L\mu_t^2-\mu_t)+1]^t\,\|\nabla F(\boldsymbol{m}_G(0))\|^2 + 2\rho L\mu_t^2\sigma^2\sum_{i=1}^{t}[2\rho(L\mu_t^2-\mu_t)+1]^{t-1}. \quad (15)$$

Using the Polyak-Lojasiewicz condition again and Assumption 1-2) provides

$$F(\boldsymbol{m}_G(t)) - F(\boldsymbol{m}_G^*) \leq [2\rho(L\mu_t^2-\mu_t)+1]^t\Theta + 2\rho L\mu_t^2\sigma^2\sum_{i=1}^{t}[2\rho(L\mu_t^2-\mu_t)+1]^{i-1}. \quad (16)$$

According to (16), the convergence bound of HoVeFL in $T$ rounds can be given by

$$F(\boldsymbol{m}_G(T)) - F(\boldsymbol{m}_G^*) \leq [2\rho(L\mu_t^2-\mu_t)+1]^T\Theta + \frac{L\mu_t\sigma^2\{[2\rho(L\mu_t^2-\mu_t)+1]^{T+1}-1\}}{L\mu_t-1}. \quad (17)$$

To ensure the convergence of HoVeFL, the learning rate satisfies $\mu_t \leq \frac{1}{L}$ since $2\rho(L\mu_t^2-\mu_t)+1 \leq 1$.

**Corollary 1.** *The convergence upper bound is convex with respect to the number of communication rounds, i.e., $T$, if $\mu_t \leq \frac{1}{L}$ and $\Theta \geq \frac{(1-2\rho\mu_t+2\rho L\mu_t^2)L\mu_t\sigma_2^2}{1-L\mu_t}$.*

*Proof.* Write the right-hand side of (17) as a function of $T$, i.e., $\mathcal{B}(T)$. By deriving the second-order derivative of $\mathcal{B}(T)$ with respect to $T$ and letting $\frac{\partial^2\mathcal{B}(T)}{\partial T^2} \geq 0$, we obtain $\mu_t \leq \frac{1}{L}$ and $\Theta \geq \frac{(1-2\rho\mu_t+2\rho L\mu_t^2)L\mu_t\sigma_2^2}{1-L\mu_t}$. □

It is important to address the difference between the update strategy in HoVeFL and the one in FedProx. While HoVeFL shares similarities in terms of employing a proximal term to stabilize the learning across non-IID data distributions, HoVeFL extends by integrating HFL and VFL. This integration is crucial in EdgeIoT, where device capabilities and data characteristics are diverse. Our convergence analysis demonstrates the theoretical underpinnings of our method.

## V. Numerical Results

In this section, we implement the proposed HoVeFL using PyTorch on a workstation with a GeForce RTX 2080 GPU.

Specific algorithmic steps of HoVeFL are as follows.

- Local Data Processing and Feature Extraction: Each device processes its data locally using a Split Convolution Neural Network with convolution layers to extract relevant features. For devices engaged in VFL, feature extraction maximizes the information gain from different data types; for HFL devices, the aim is to harmonize feature representation across non-IID datasets.
- Local Model Update: The local loss function in (1) is minimized, leading to an optimal local model $\boldsymbol{m}_{i,j}^{n*}(t)$ at device $n$. Local model updates can be performed using a specified learning algorithm (e.g., SGD for handling data heterogeneity).
- Model Aggregation: At the server, the model aggregation is conducted to combine vertical and horizontal updates.
- Global Model Synthesis and Distribution: After aggregation, a weighted model aggregation aggregates hybrid local models so that a global loss function $F(\boldsymbol{m}_G(t))$ in (4) is minimized. The global model $\boldsymbol{m}_G(t)$ is distributed across devices for further local training and refinement.

Figs. 2 and 3 show the training and testing losses of HoVeFL on CIFAR-10 and street view house numbers (SVHN) datasets, respectively. In particular, the CIFAR-10 and SVHN datasets are non-specialized but well-recognized benchmark datasets for evaluating the performance of FL frameworks. CIFAR-10 contains 60,000 color images of size 32×32, categorized into 10 classes (e.g., airplanes, cats, trucks), providing a diverse set of small, natural objects. SVHN consists of over 600,000 digit images captured from real-world street views, focusing on numbers in various positions and lighting conditions. The distinct nature of these datasets, i.e., objects versus digits, introduces variability in terms of content, image distribution, and classification tasks. In addition, we distribute the data in a non-IID fashion among HFL devices to reflect real-world variability. Adam and SGD optimizers were tested on HoVeFL. We also adapted the numbers of devices performing HFL (i.e., $\mathbf{N_i}$) and devices performing VFL (i.e., $\mathbf{N_j}$). The performance of HoVeFL is compared with the one where all $N$ devices conduct VFL.

Given 12 HFL devices and 6 VFL devices in HoVeFL, the training and testing losses with the Adam optimizer converges much faster than those with the SGD optimizer, confirming the convergence of Adam and SGD in [15]. This validates the effectiveness of HoVeFL with heterogeneous HFL and VFL devices. Given the Adam optimizer, the testing loss of HoVeFL with 12 HFL devices and 6 VFL devices is 5.5% and 25.2% higher than the one with 6 HFL devices and 12 VFL devices on the CIFAR-10 and SVHN datasets, respectively. This is because the training dataset at the VFL devices ($N_j$) in HoVeFL is IID (due to training the local models on similar data samples $j$ but for different features) while the data at the HFL devices ($N_i$) in HoVeFL is non-IID (due to the same data feature $i$ but on different data samples).

Given the CIFAR-10 and SVHN datasets, the testing losses of HoVeFL (Adam, H6:V12) are 5.9% and 33.3% higher than those of VFL (Adam), respectively. This is reasonable since the devices carrying out HFL in HoVeFL hold data with the same feature space but distinct and potentially non-IID data samples. Moreover, the testing losses of HFL (Adam) are 19% higher than those of HoVeFL (Adam, H12:H6). This is because a combination of different data samples from VFL devices and different features from HFL devices helps the model generalize across different characteristics and types of data. The proposed HoVeFL minimizes the loss function effectively, leading to mild testing losses.

In addition, Figs. 2 and 3 show that the training and testing

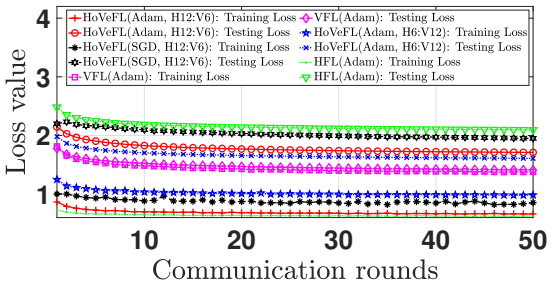


Fig. 2. The training and testing loss based on CIFAR-10.

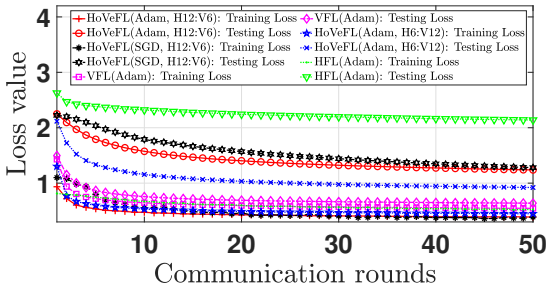


Fig. 3. The training and testing loss based on SVHN.

losses of HoVeFL have a significant difference, compared to VFL. This is because the training of HoVeFL includes a wide variety of data types and distributions, leading to difficulties in achieving a model that generalizes across all types of data. The model can capture specific features or patterns present in the training data but not representative of the general population. Given new or unseen data in the test set, the goal of generalizing across different data features and populations makes HoVeFL more susceptible to discrepancies in performance. On the other hand, VFL focuses on vertically integrating features for the same set of data samples. The diversity in data types is less pronounced in VFL than it is in HoVeFL. The alignment of features across the same set of the clients in VFL ensures that training and testing datasets are likely to follow similar distributions, thus maintaining a closer performance metric between training and testing.

## VI. Conclusions and Future Work

In this letter, we proposed a new HoVeFL for EdgeIoT, which enables some devices to train local models on different, independent features of the same data samples, while the other devices train local models for the same features using distinct, non-overlapping data samples. The training and testing losses were evaluated on the CIFAR-10 and SVHN datasets, with Adam and SGD. We also experimentally varied the number of devices performing HFL and VFL in HoVeFL. The performance was compared with a scenario where all EdgeIoT devices conduct VFL.

In future work, we intend to apply the proposed HoVeFL framework to real-world medical datasets as they become accessible. Access to such data will allow us to validate and refine our approach within actual healthcare scenarios, addressing the unique challenges and requirements of medical data analysis. By testing HoVeFL with authentic medical datasets, we aim to demonstrate its practical effectiveness and its potential to enhance collaborative learning while preserving patient privacy in healthcare environments.

## Acknowledgements

This work was supported by the CISTER Research Unit (UIDP/UIDB/04234/2020) and project ADANET (PTDC/EEICOM/3362/2021), financed by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology); and also supported in part by the AXA Research Fund (AXA Chair for Internet of Everything at Koç University).

## References

[1] K. Li, Y. Cui, W. Li, T. Lv, X. Yuan, S. Li, W. Ni, M. Simsek, and F. Dressler, "When internet of things meets metaverse: Convergence of physical and cyber worlds," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4148–4173, March 2022.

[2] L. You, S. Liu, B. Zuo, C. Yuen, D. Niyato, and H. V. Poor, "Federated and asynchronized learning for autonomous and intelligent things," *IEEE Network*, vol. 38, no. 2, pp. 286–293, March 2023.

[3] J. Park, J. Moon, T. Kim, P. Wu, T. Imbiriba, P. Closas, and S. Kim, "Federated learning for indoor localization via model reliability with dropout," *IEEE Communications Letters*, vol. 26, no. 7, pp. 1553–1557, July 2022.

[4] P. Liu, G. Zhu, W. Jiang, W. Luo, J. Xu, and S. Cui, "Vertical federated edge learning with distributed integrated sensing and communication," *IEEE Communications Letters*, vol. 26, no. 9, pp. 2091–2095, September 2022.

[5] H. Liu, X. Zhang, X. Shen, H. Sun, and M. Shahidehpour, "A hybrid federated learning framework with dynamic task allocation for multi-party distributed load prediction," *IEEE Transactions on Smart Grid*, vol. 14, no. 3, pp. 2460–2472, May 2022.

[6] X. Zhang, W. Yin, M. Hong, and T. Chen, "Hybrid federated learning: Algorithms and implementation," in *NeurIPS Workshop on Scalability, Privacy, and Security in Federated Learning (SpicyFL)*, December 2020.

[7] J. L. Salmeron and I. Arévalo, "Concurrent vertical and horizontal federated learning with fuzzy cognitive maps," *Future Generation Computer Systems*, vol. 162, p. 107482, January 2025.

[8] C. Thapa, P. C. M. Arachchige, S. Camtepe, and L. Sun, "Splitfed: When federated learning meets split learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8485–8493.

[9] X. Liu, Y. Deng, and T. Mahmoodi, "Wireless distributed learning: A new hybrid split and federated learning approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2650–2665, April 2023.

[10] K. Li, J. Zheng, X. Yuan, W. Ni, O. B. Akan, and H. V. Poor, "Data-agnostic model poisoning against federated learning: A graph autoencoder approach," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 3465–3480, February 2024.

[11] J. Zheng, K. Li, N. Mhaisen, W. Ni, E. Tovar, and M. Guizani, "Exploring deep-reinforcement-learning-assisted federated learning for online resource allocation in privacy-preserving edgeiot," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 099–21 110, November 2022.

[12] K. Li, J. Zheng, W. Ni, H. Huang, P. Liò, F. Dressler, and O. B. Akan, "Biasing federated learning with a new adversarial graph attention network," *IEEE Transactions on Mobile Computing*, 2024.

[13] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16*. Springer, 2016, pp. 795–811.

[14] A. Elgabli, C. B. Issaid, A. S. Bedi, K. Rajawat, M. Bennis, and V. Aggarwal, "Fednew: A communication-efficient and privacy-preserving newton-type method for federated learning," in *International Conference on Machine Learning (ICML)*. PMLR, 2022, pp. 5861–5877.

[15] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, "The marginal value of adaptive gradient methods in machine learning," *Advances in neural information processing systems*, vol. 30, 2017.