

Adaptive Federated Learning for Future IoV-Oriented IoT End-to-End Network Planning

Jiaming Pei, *Member, IEEE*, Xiaoqing Xu, Lukun Wang, Saba Al-Rubaye, Sun Zhang and Anwer Al-Dulaimi, *Senior Member, IEEE*

Abstract—In the Internet of Things (IoT) domain, end-to-end (E2E) planning tasks require distributed devices to collaboratively train deep models under highly dynamic environments. However, existing federated learning (FL) methods often assume homogeneous communication conditions and static node reliability, leading to suboptimal aggregation performance when confronted with heterogeneous uncertainty sources such as sensing noise, prediction bias, and communication instability. To address this challenge, we propose FedUPAP (Federated Uncertainty-Aware End-to-End Planning), a novel framework that dynamically adjusts client contributions based on multi-source uncertainty and network topology information. Specifically, each IoV vehicle node within the broader IoT system estimates three uncertainty factors—prediction uncertainty, sensing uncertainty, and communication uncertainty—to represent its model reliability and transmission stability. A topology-aware weighting module further refines the aggregation by incorporating node connectivity and link quality. In addition, a temporal smoothing strategy is introduced to stabilize weight evolution over successive communication rounds. Extensive experiments on various E2E IoV-centric IoT planning scenarios demonstrate that FedUPAP achieves superior convergence stability, communication efficiency, and planning accuracy compared with existing adaptive aggregation and uncertainty-based FL baselines. The proposed approach provides a promising direction toward uncertainty-robust and topology-adaptive federated optimization in large-scale IoT and IoV networks.

Index Terms—End-to-end planning, federated learning, uncertainty-aware aggregation.

I. INTRODUCTION

WITH the rapid evolution of the Internet of Vehicles [1] (IoV) and intelligent transportation systems [2] (ITS), end-to-end [3] (E2E) planning has become a pivotal paradigm for autonomous driving. Unlike traditional modular pipelines—which separately handle perception, prediction, decision-making, and control—E2E approaches use deep neural networks [4] (DNNs) to directly map sensory inputs to

This work was supported by the Humanity and Social Science Research Project of Anhui Educational Committee (Grant No. 2024AH052138) and the University-Industry Collaborative Education Program of the Ministry of Education (Grant No. 231100751213206).

Jiaming Pei is with the School of Computer Science, University of Sydney, Sydney, NSW 2006, Australia. (e-mail: jpei0906@uni.sydney.edu.au)

Lukun Wang and Xiaoqing Xu are with the School of Intelligent Equipment, Shandong University of Science and Technology, Tai'an, Shandong, 271000, China. email:202383230053,wanglukun@sdust.edu.cn.

Saba Al-Rubaye is with the Faculty of Engineering and Applied Science, Cranfield University. (e-mail: S.Alrubaye@cranfield.ac.uk.)

Sun Zhang is with the Department of Computer Science and Technology, Anhui University of Finance and Economics, Bengbu, China (e-mail: sun-zhang@aufe.edu.cn). (Corresponding author)

Anwer Al-Dulaimi is with the College of Technological Innovations, Zayed University, Abu Dhabi, United Arab Emirates (e-mail: anwer.aldulaimi@ieee.org)

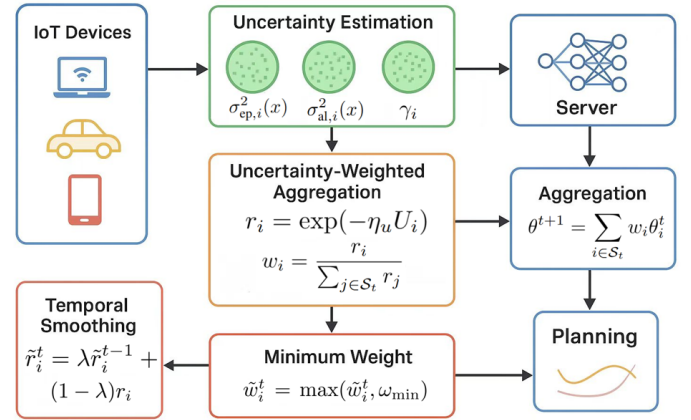


Fig. 1. FedUPAP: Federated Uncertainty-Aware End-to-End Planning

driving actions [5], enhancing model expressiveness and global optimization for large-scale IoT-driven vehicular systems.

However, E2E performance heavily depends on training data diversity and consistency. In real-world IoV environments, vehicles operate under varying weather, lighting, traffic, and sensor reliability. Centralized training is often infeasible due to privacy, communication overhead, and data drift. To address this, federated learning (FL) enables distributed model training across multiple vehicles or edge devices [6]–[8], exchanging only model parameters instead of raw data. This preserves data locality and security while supporting global aggregation, and has shown potential in vehicular networks, industrial IoT (IIoT [9]), and medical IoT (IoMT [10]).

Directly applying FL to E2E planning faces challenges. Conventional aggregation algorithms—FedAvg [11], FedProx [12], FedNova [13], and FedDyn [14]—assume equal contribution or static importance weights. In heterogeneous IoV environments, nodes differ in data quality, environmental complexity, and training stability, leading to global model degradation, unstable decision-making, and reduced robustness. The inability to dynamically adjust aggregation importance based on node uncertainty is a key bottleneck in federated E2E planning.

Recent studies attempt to address this in three directions. First, personalized FL methods [15] enhance individual adaptability via local layers or attention [16] but ignore node reliability in aggregation. Second, adaptive weighted aggregation methods [17] assign weights based on loss or gradient variance, reflecting only surface-level dynamics and failing to capture predictive confidence. Third, Bayesian and uncertainty-

aware FL models predictive uncertainty but incurs heavy computational and communication costs, limiting scalability in E2E tasks.

To overcome these challenges, this paper proposes FedUAP, introducing Uncertainty-Weighted Aggregation (UWA) to dynamically adjust node importance based on epistemic and aleatoric uncertainties. We theoretically prove that, under mild assumptions on bounded predictive uncertainty and Lipschitz continuity of local loss, UWA ensures a tighter upper bound on global loss than uniform aggregation, guaranteeing improved convergence and robustness. Let $\mathcal{L}(\theta)$ be the expected global loss, and w_i the uncertainty-weighted contribution of client i :

$$\mathcal{L}(\theta^{t+1}) \leq \sum_{i \in S_t} w_i \mathcal{L}_i(\theta_i^t) + \epsilon, \quad (1)$$

where ϵ bounds approximation errors from stochastic local updates. This shows nodes with lower predictive uncertainty influence the global model more, improving reliability.

Each client estimates local uncertainty using a lightweight module. The server assigns aggregation weights based on predictive confidence, favoring reliable nodes. This theoretical analysis supports the claimed advantages of FedUAP in Section I, ensuring stable and robust convergence under heterogeneous IoV conditions.

FedUAP also incorporates an uncertainty propagation and feedback channel across perception–planning stages, improving interpretability and robustness without increasing communication or computation cost.

The main contributions are:

- Problem formulation and framework design: Identify static importance limitations and propose uncertainty-aware FedUAP.
- Uncertainty-Weighted Aggregation (UWA) algorithm: Jointly integrates epistemic and aleatoric uncertainty for adaptive model fusion.
- End-to-end uncertainty propagation: Enables uncertainty flow across stages, improving consistency and interpretability.
- Extensive experiments and analysis: Demonstrates superior trajectory accuracy, convergence stability, and robustness across IoT scenarios.

II. RELATED WORK

Research on end-to-end (E2E) planning combined with federated learning can be broadly categorized into centralized E2E planning [18], federated E2E planning [19], uncertainty modeling [20], and federated adaptive aggregation methods. Centralized E2E planning approaches aim to directly map sensory inputs to vehicle control outputs [21] through unified training. Early methods employed convolutional neural networks (CNNs) to process forward-facing camera images and predict steering angles or velocities, which achieved good performance in limited scenarios but lacked adaptability to complex environments. More recent works incorporate multi-modal inputs, including camera images, LiDAR point clouds, radar, and V2X information [22]. Features from different

modalities are fused through concatenation, weighted attention mechanisms, or Transformer encoders to capture spatial-temporal correlations. Some methods further adopt graph convolutional networks or global attention modules to model interactions with surrounding traffic participants [23]. Multi-task learning frameworks are also leveraged to jointly optimize trajectory prediction, velocity planning, and behavior prediction, aiming to improve generalization through shared low-level features. Recent studies have explored federated learning algorithms that adapt aggregation weights dynamically via optimal transport [24] and Bayesian multi-task aggregation under heterogeneous preferences [25], highlighting a trend toward more adaptive and theoretically grounded aggregation schemes. Despite these advances, centralized E2E methods remain heavily dependent on large-scale labeled datasets and are sensitive to environmental and sensor variations. When deployed across different cities, road conditions, or sensor configurations, their performance degrades rapidly. More importantly, centralized models cannot adapt to heterogeneous nodes in distributed IoT environments, and they are incapable of dynamically reflecting each node’s contribution to the global model, highlighting their limitation for federated scenarios.

To address data isolation, federated learning has been integrated into E2E planning, enabling vehicles or edge nodes to collaboratively train a global model without sharing raw data. Representative approaches include FedUAP aggregation with periodic updates, local fine-tuning to improve global performance, and task-similarity-based weighted aggregation. Some studies incorporate meta-learning for personalized adaptation or knowledge distillation to align local models with the global model [26], [27]. In addition, very recent work has studied Bayesian federated learning for multiple tasks simultaneously, combining classification and regression under different client objectives (“Task Diversity in Bayesian Federated Learning”) [28]. However, these methods often rely on static or delayed aggregation weights, which fail to capture the actual reliability of nodes during training. Nodes with limited data, high noise, or complex environments may still exert disproportionate influence on the global model, reducing convergence speed and trajectory prediction accuracy. Even when task similarity or historical performance is used for weighting, these metrics often cannot reflect the node’s real-time credibility in dynamic environments. This limitation directly motivates the need for our proposed approach: dynamically adjusting node contributions based on predictive uncertainty during aggregation.

Uncertainty modeling has been widely adopted in E2E planning to enhance decision-making safety and model robustness [29]. Common techniques include Bayesian neural networks, Monte Carlo Dropout, and heteroscedastic regression [30]. In FedUAP, predictive uncertainty estimation is integrated into the aggregation process, allowing the global model to weigh contributions based on confidence. Unlike previous approaches, this mechanism is specifically tailored for E2E decision-making, ensuring practical scalability in large-scale IoT networks. Moreover, recent federated aggregation research has proposed learnable aggregation weights with resilience to Byzantine faults (“Byzantine-Robust Federated Learning with Learnable Aggregation Weights”) [31], which further

validates the feasibility of adaptive weight schemes. These techniques allow risk-aware decision-making, adaptive loss weighting, and confidence estimation, while also supporting dynamic aggregation in heterogeneous settings.

Federated adaptive aggregation methods have been extensively studied to mitigate the impact of node heterogeneity [32]–[35], reflecting a well-established research direction in FL. In FedUAP, the Uncertainty-Weighted Aggregation (UWA) mechanism builds upon these prior works by jointly considering predictive uncertainty when assigning aggregation weights, providing more fine-grained control over node contributions. While attention-based approaches, knowledge distillation, and optimization-based strategies can be viewed as specific instances within this broader aggregation design, FedUAP explicitly models node reliability to further enhance performance. Moreover, very recent studies (2025) have proposed aggregation via optimal transport to exactly align client distributions (FedAVOT) [36], offering a theoretically principled foundation for weight design in heterogeneous FL. By integrating uncertainty-aware weighting with these adaptive aggregation principles, high-uncertainty nodes contribute less to global updates, thereby improving stability and trajectory prediction accuracy in E2E planning.

Centralized E2E planning lacks distributed adaptability, conventional federated E2E approaches employ static aggregation weights, uncertainty modeling is rarely integrated into global aggregation, and adaptive aggregation strategies often overlook nodes' dynamic reliability. These limitations collectively highlight the core problem addressed in this work: in federated E2E planning, it is crucial to dynamically adjust node contributions during aggregation based on predictive uncertainty. The proposed FedUAP framework fulfills this need by incorporating node-level uncertainty modeling and adaptive weighting, thereby improving multi-node collaboration, planning accuracy, and robustness in heterogeneous IoT environments.

III. METHODOLOGY

In this section, we present the detailed design and implementation of the proposed FedUAP (Federated Uncertainty-Aware End-to-End Planning) framework (As shown in Fig.1). The overall goal of FedUAP is to dynamically regulate the contribution of heterogeneous IoT nodes during federated aggregation, thereby mitigating the performance degradation observed in traditional methods such as FedAvg [11] under high heterogeneity. Specifically, FedUAP explicitly models multi-source uncertainties at each client and integrates them into a dynamic weighting mechanism, while incorporating temporal smoothing and end-to-end feedback to enhance planning stability and robustness.

Throughout this section, $\mathcal{S}_t$ denotes the set of participating IoT clients selected at communication round t . Unless otherwise specified, all client indices i and j appearing in the following equations satisfy $i \in \mathcal{S}_t$ and $j \in \mathcal{S}_t$, respectively.

The client-side uncertainty estimation forms the basis for assessing node reliability. We decompose uncertainty into two principal types: epistemic uncertainty, capturing the model's

Algorithm 1 FedUAP: Federated Uncertainty-Aware End-to-End Planning

Require: Global model θ^0 , total communication rounds T , local epochs E , learning rate η , smoothing factor λ , minimum weight $\omega_{\min}$, temperature η_u

- 1: **for** round $t = 0$ **to** $T - 1$ **do**
- 2: Server selects a subset of clients $\mathcal{S}_t$
- 3: Server broadcasts global model θ^t to all $i \in \mathcal{S}_t$
- 4: **for** each client $i \in \mathcal{S}_t$ in parallel **do**
- 5: Initialize local model $\theta_{i,0} = \theta^t$
- 6: **for** local epoch $k = 0$ **to** $E - 1$ **do**
- 7: Compute local loss $\mathcal{L}_i^{\text{local}}(\theta_{i,k})$ using heteroscedastic regression
- 8: Update local model: $\theta_{i,k+1} = \theta_{i,k} - \eta \nabla_{\theta} \mathcal{L}_i^{\text{local}}(\theta_{i,k})$
- 9: **end for**
- 10: Estimate local epistemic uncertainty $\bar{\sigma}_{\text{ep},i}^2$ via Monte Carlo Dropout
- 11: Estimate local aleatoric uncertainty $\bar{\sigma}_{\text{al},i}^2$
- 12: Compute communication reliability γ_i
- 13: Compute combined uncertainty: $U_i = \alpha \bar{\sigma}_{\text{ep},i}^2 + \beta \bar{\sigma}_{\text{al},i}^2 + \gamma(1 - \gamma_i)$
- 14: Send $(\theta_i^t, U_i, \gamma_i)$ to server
- 15: **end for**
- 16: **for** each client $i \in \mathcal{S}_t$ **do**
- 17: Compute raw weight: $r_i = \exp(-\eta_u U_i)$
- 18: Smooth weight: $\tilde{r}_i^t = \lambda \tilde{r}_i^{t-1} + (1 - \lambda) r_i$
- 19: **end for**
- 20: Normalize weights: $\tilde{w}_i^t = \frac{\tilde{r}_i^t}{\sum_{j \in \mathcal{S}_t} \tilde{r}_j^t}$
- 21: Apply minimum weight bound: $\tilde{w}_i^t = \max(\tilde{w}_i^t, \omega_{\min})$ and re-normalize
- 22: Aggregate global model: $\theta^{t+1} = \sum_{i \in \mathcal{S}_t} \tilde{w}_i^t \theta_i^t$
- 23: **end for**

Ensure: Final global model θ^T

confidence on a given input or distribution, and aleatoric uncertainty, representing inherent observation noise or environmental variability. Technically, each client network introduces two output branches: one for the predicted action or trajectory $\hat{y}(x)$ and another for log-variance estimation $\hat{s}(x)$. Epistemic uncertainty is approximated via Monte Carlo Dropout (or lightweight ensemble) using M forward passes, yielding a set of stochastic predictions $\{\hat{y}^{(m)}(x)\}_{m=1}^M$. The sample mean and variance are then used to estimate the predictive distribution:

$$\mu_i(x) = \frac{1}{M} \sum_{m=1}^M \hat{y}_i^{(m)}(x), \quad (2)$$

$$\sigma_{\text{ep},i}^2(x) = \frac{1}{M} \sum_{m=1}^M (\hat{y}_i^{(m)}(x) - \mu_i(x))^2, \quad (3)$$

where $\hat{y}_i^{(m)}(x)$ is the m -th stochastic forward pass of client i , and M is the number of samples. The aleatoric uncertainty is obtained by exponentiating the log-variance branch:

$$\sigma_{\text{al},i}^2(x) = \exp(\hat{s}_i(x)), \quad (4)$$

where $\hat{s}_i(x)$ is the predicted log-variance, and exponentiation is used to ensure the aleatoric variance is positive. This follows standard heteroscedastic regression practice and allows stable training while modeling observation noise.

$$\mathcal{L}_i^{\text{local}}(\theta) = \mathbb{E}_{(x,y) \sim \mathcal{D}_i} \left[\frac{1}{2} \exp(-\hat{s}_i(x)) \|y - \hat{y}_i(x)\|^2 + \frac{1}{2} \hat{s}_i(x) \right] + \rho \|\theta - \theta^t\|^2, \quad (5)$$

This local objective is designed based on heteroscedastic regression principles. The first term models observation noise, allowing the network to weigh samples inversely proportional to their variance. The proximal term $\rho \|\theta - \theta^t\|^2$ ensures stability under heterogeneous nodes, encouraging local updates to stay close to the global model. This design provides both theoretical justification and practical stability for federated aggregation.

The loss serves to reduce the influence of high-noise samples while retaining predictive uncertainty estimates for subsequent aggregation. This explicitly ensures that nodes with noisy data do not dominate the aggregation, while their uncertainty information is preserved for the UWA mechanism.

In addition to prediction and observation uncertainty, FedUAP incorporates communication uncertainty to reflect the reliability of parameter transmission. We define a communication reliability metric γ_i based on round-trip delay d_i , packet loss p_i , and recent upload success ratio q_i :

$$\gamma_i = \exp \left(-\eta_d \frac{d_i}{d_{\max}} - \eta_p \frac{p_i}{p_{\max}} \right) \cdot \frac{q_i}{q_{\min}}, \quad (6)$$

where $d_{\max}$, $p_{\max}$, and $q_{\min}$ are normalization constants chosen based on empirical measurements (e.g., 95th percentile for delay and loss, minimum expected success ratio). This normalization ensures that the three types of uncertainty are on comparable scales, preventing domination of any single factor in the aggregation weight.

The combined node uncertainty is then computed as:

$$U_i = \alpha \bar{\sigma}_{\text{ep},i}^2 + \beta \bar{\sigma}_{\text{al},i}^2 + \gamma(1 - \gamma_i), \quad (7)$$

where α, β, γ are chosen such that the normalized uncertainties contribute proportionally to the aggregation weight.

On the server, we apply Uncertainty-Weighted Aggregation (UWA) to compute aggregation weights from the uncertainties:

$$r_i = \exp(-\eta_u U_i), \quad w_i = \frac{r_i}{\sum_{j \in \mathcal{S}_t} r_j}, \quad (8)$$

Here, the summation index j is explicitly restricted to the same participating client set $\mathcal{S}_t$ to ensure consistent notation and avoid ambiguity.

To further improve aggregation under heterogeneous network topologies, we incorporate a topology-aware module that adjusts node importance based on connectivity and update freshness. The module computes a topology-aware score for each node i as

$$w_i^{\text{topo}} = \frac{f(s_i, d_i, c_i)}{\sum_{j \in \mathcal{S}_t} f(s_j, d_j, c_j)} \quad (9)$$

where s_i is the staleness of local updates, d_i is the node degree in the network graph, c_i represents communication quality, and $f(\cdot)$ combines these factors into a single score. The final aggregation weight is then

$$w_i^{\text{final}} = w_i \cdot w_i^{\text{topo}}, \quad \tilde{w}_i^t = \frac{w_i^{\text{final}}}{\sum_{j \in \mathcal{S}_t} w_j^{\text{final}}} \quad (10)$$

This ensures that nodes with strong connectivity, reliable communication, and fresh updates contribute more, while poorly connected or outdated nodes are down-weighted. The computation is lightweight and performed on the server each round, introducing negligible latency while improving convergence stability.

The global model update is then

$$\theta^{t+1} = \sum_{i \in \mathcal{S}_t} w_i^{\text{final}} \theta_i^t, \quad (11)$$

with θ_i^t being the locally updated parameters. To further stabilize aggregation, an exponential weighted moving average (EWMA) smooths the weights across rounds:

$$\tilde{r}_i^t = \lambda \tilde{r}_i^{t-1} + (1 - \lambda) r_i, \quad \tilde{w}_i^t = \frac{\tilde{r}_i^t}{\sum_{j \in \mathcal{S}_t} \tilde{r}_j^t}, \quad (12)$$

where $\lambda \in [0, 1]$ is the smoothing factor.

$$\mathcal{R}_{\text{fair}}(\mathbf{w}) = \eta_f \sum_{i \in \mathcal{S}_t} \left(\tilde{w}_i^t - \frac{1}{|\mathcal{S}_t|} \right)^2 \quad (13)$$

This regularization term is included to ensure long-term participation of all clients.

Finally, FedUAP introduces an end-to-end uncertainty feedback channel, enabling the planning module to adjust trajectory or control decisions based on predicted uncertainties. High uncertainty triggers conservative planning or safety-aware control adjustments, enhancing decision robustness without additional model complexity. The overall algorithm proceeds as follows: in round t , the server broadcasts θ^t ; each client $i \in \mathcal{S}_t$ performs E steps of local SGD

$$\theta_{i,k+1} = \theta_{i,k} - \eta \nabla_{\theta} \mathcal{L}_i^{\text{local}}(\theta_{i,k}), \quad k = 0, \dots, E-1, \quad \theta_{i,0} = \theta^t, \quad (14)$$

Here, η denotes the learning rate that controls the step size of each local update.

Then, the client computes U_i , sends $(\theta_i^t, U_i, \gamma_i)$ to the server, which calculates smoothed weights $\tilde{w}_i^t$ and updates θ^{t+1} . Communication overhead is minimal, as only a few scalars (U_i and γ_i) are transmitted per round.

The computational complexity of FedUAP primarily stems from three components: the local model training, uncertainty estimation, and the aggregation phase. The complexity for each client's local model training is $O(E \cdot D \cdot F)$, while uncertainty estimation adds $O(M \cdot D)$. Aggregation complexity is $O(N)$. Therefore, the overall complexity per round is $O(N \cdot (E \cdot D \cdot F + M \cdot D))$.

In terms of scalability, the proposed framework is designed to handle increasing node heterogeneity. The communication

overhead is minimal, making it highly scalable for large IoT networks.

Overall, FedUAP systematically integrates multi-source uncertainty estimation, uncertainty-weighted aggregation, temporal smoothing, fairness constraints, topology-aware weighting, and end-to-end feedback to achieve robust, adaptive federated E2E planning.

IV. EXPERIMENTS

To evaluate the effectiveness and robustness of FedUAP in end-to-end IoT planning scenarios, we design a series of experiments to assess its performance under node heterogeneity, varying communication conditions, and uncertainty. The experiments focus on three aspects: global model convergence stability, end-to-end planning accuracy, and communication efficiency. By comparing FedUAP with several existing federated learning methods, we analyze its advantages in handling high-uncertainty nodes and demonstrate its scalability across diverse IoT scenarios. All experiments are conducted under a unified hardware and software environment to ensure reproducibility and fairness.

TABLE I
DATASET INFORMATION

Dataset	Devices / Data Sources	Application Domain
FedAIoT	Various smart devices	Federated learning benchmark
CIC IoT 2023	105 IoT devices	IoT security and intrusion detection
TON_IoT	Industrial IoT devices	Industrial IoT security and planning
IoT-23	IoT devices	IoT network traffic and anomaly detection
Multivariate IoT & AI	27 global cities	Smart city development indicators
Time Series in IoT	Traffic sensors	Traffic pattern prediction

A. Datasets and Experimental Setup

Six publicly available and real-world datasets related to end-to-end IoT planning are used to simulate collaborative training across different IoT nodes and application scenarios. These include FedAIoT (comprehensive smart device data for evaluating federated learning performance), CIC IoT 2023 (network traffic from 105 IoT devices for security and intrusion detection tasks), TON_IoT (industrial IoT data for industrial environment planning and security studies), IoT-23 (IoT network traffic and anomaly detection data), Multivariate IoT & AI (data from 27 global cities for smart city development indicators), and Time Series in IoT (traffic sensor time-series data for traffic pattern prediction). These datasets cover a wide range of scenarios from smart homes and wearable devices to industrial IoT and urban traffic management, providing a comprehensive testbed for evaluating FedUAP under different node heterogeneity and environmental conditions. Key dataset information is summarized in Table I.

To rigorously validate FedUAP, we design controlled experimental scenarios on these datasets to simulate heterogeneous IoT nodes with varying data distributions, sample

TABLE II
COMPARISON OF SIX ALGORITHMS ON SIX IoT DATASETS. ROUNDS REPRESENTS ROUNDS TO 75% ACCURACY AND VARIANCE REPRESENTS HETEROGENEOUS NODE VARIANCE.

FedAIoT				
Algorithm	Rounds	ACC	Variance	Stability
FedUAP	18	0.96	0.002	0.004
FedAvg	36	0.92	0.008	0.010
FedProx	30	0.94	0.006	0.008
FedMA	40	0.90	0.010	0.012
FedDyn	28	0.95	0.005	0.007
FedDF	32	0.93	0.007	0.009
CIC IoT 2023				
Algorithm	Rounds	ACC	Variance	Stability
FedUAP	20	0.95	0.002	0.004
FedAvg	38	0.91	0.009	0.010
FedProx	32	0.93	0.006	0.008
FedMA	42	0.89	0.011	0.013
FedDyn	30	0.94	0.005	0.007
FedDF	34	0.92	0.008	0.009
TON_IoT				
Algorithm	Rounds	ACC	Variance	Stability
FedUAP	22	0.94	0.003	0.004
FedAvg	40	0.90	0.010	0.011
FedProx	34	0.92	0.007	0.009
FedMA	44	0.88	0.012	0.013
FedDyn	32	0.93	0.006	0.008
FedDF	36	0.91	0.009	0.010
IoT-23				
Algorithm	Rounds	ACC	Variance	Stability
FedUAP	25	0.92	0.003	0.004
FedAvg	42	0.88	0.011	0.012
FedProx	36	0.90	0.007	0.009
FedMA	46	0.86	0.013	0.014
FedDyn	34	0.91	0.006	0.008
FedDF	38	0.89	0.009	0.010
Multivariate IoT & AI				
Algorithm	Rounds	ACC	Variance	Stability
FedUAP	28	0.90	0.004	0.005
FedAvg	45	0.85	0.012	0.013
FedProx	38	0.88	0.008	0.010
FedMA	50	0.83	0.014	0.015
FedDyn	36	0.89	0.007	0.009
FedDF	40	0.87	0.010	0.011
Time Series in IoT				
Algorithm	Rounds	ACC	Variance	Stability
FedUAP	26	0.91	0.003	0.004
FedAvg	43	0.86	0.011	0.012
FedProx	37	0.89	0.007	0.009
FedMA	48	0.84	0.013	0.014
FedDyn	35	0.90	0.006	0.008
FedDF	39	0.88	0.009	0.010

sizes, and noise levels. Communication constraints such as latency and packet loss are emulated to reflect real-world network conditions. This ensures that the experimental setup goes beyond simple dataset usage and allows us to assess FedUAP's performance under realistic collaborative learning environments.

To fully evaluate FedUAP, five representative end-to-end IoT federated learning algorithms are selected for comparison. FedAvg [11] serves as the basic model averaging method, suitable for scenarios with relatively homogeneous node data and stable communication but may struggle with convergence under non-IID data. FedProx [12] extends FedAvg by adding

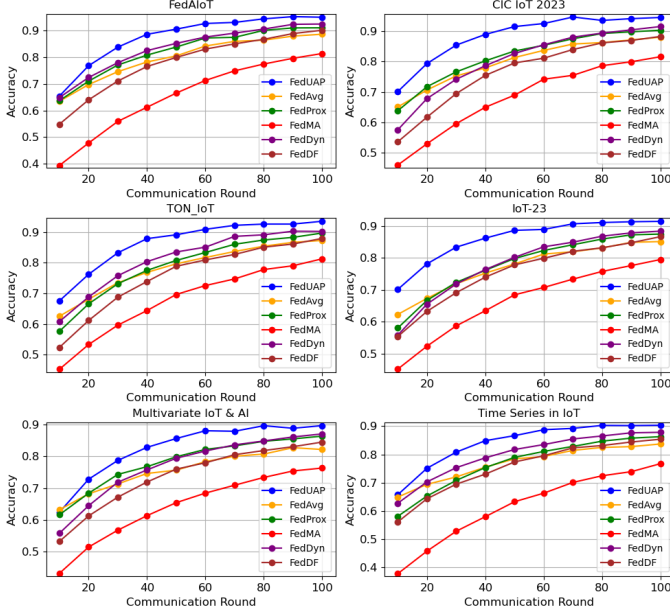


Fig. 2. Performance comparison of six federated learning algorithms on six IoT datasets. The algorithms are: FedUAP (blue), FedAvg (orange), FedProx (green), FedMA (red), FedDyn (purple), and FedDF (brown).

a proximal term to mitigate the impact of heterogeneous data on the global model, improving convergence stability. FedMA [37] aligns client model weights before aggregation to enhance compatibility under non-IID data and differing model architectures. FedDyn [14] dynamically adjusts aggregation strategies based on client data distribution and local loss variations, enhancing convergence and stability in dynamic node environments. FedDF [38] integrates client model knowledge via knowledge distillation, improving global model generalization when client models differ significantly or have uneven data distributions. Selecting these algorithms provides a diverse set of aggregation strategies and allows for a comprehensive demonstration of FedUAP’s advantages in dynamic weighting and uncertainty handling.

All experiments are conducted under a unified hardware and software environment. The hardware consists of an Intel Core i7-10700K CPU, 32GB of RAM, and an NVIDIA RTX 3070 GPU. The software environment includes Python 3.9, PyTorch 1.10, and CUDA 11.3, with the Flower v0.16.0 federated learning framework. The optimizer is Adam with a learning rate of 0.001, batch size of 32, 10 local training epochs per round, 10 clients selected per communication round, and a total of 50 communication rounds. This controlled and unified experimental setup ensures fairness across all methods and provides sufficient rigor to validate FedUAP’s effectiveness across diverse IoT scenarios with heterogeneous nodes and dynamic conditions.

Additionally, the proposed FedUAP framework introduces minimal computational overhead. The main computational cost lies in the uncertainty estimation at each client, which involves M forward passes per client (for epistemic uncertainty) and log-variance estimation for aleatoric uncertainty. These operations are computationally feasible as they only

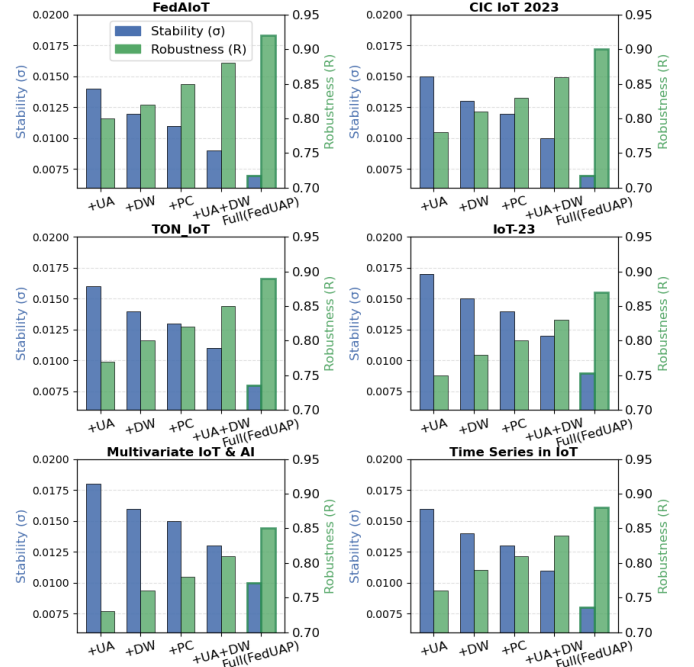


Fig. 3. Ablation study results on six IoT datasets. The left bars indicate Stability (σ), calculated as the standard deviation of accuracy in the last training rounds, reflecting the smoothness of convergence. The right bars indicate Robustness (R), calculated as the ACC achieved by the model. The Full (FedUAP) model consistently achieves the highest robustness and lowest fluctuations, demonstrating the effectiveness of dynamic weighting, contribution compensation, and adaptive stability modules.

require a few additional layers in the client network, which can be implemented efficiently with modern hardware. While the increased complexity in node aggregation is apparent due to the dynamic weighting mechanism, it remains linear with respect to the number of clients, ensuring that FedUAP is scalable in real-world IoT environments. The communication overhead is minimal as only scalar values for uncertainties and model updates are transmitted, significantly reducing the data transfer load compared to conventional methods. The scalability of FedUAP is thus confirmed for large-scale deployments, as the added complexity and overhead do not substantially hinder its practical use in large IoT networks.

B. Performance Analysis

To comprehensively evaluate the performance of FedUAP in IoT environments, this section compares four key metrics: the number of rounds required to reach 75% accuracy, ACC, heterogeneous node variance, and convergence stability. These metrics jointly assess the algorithm’s ability to address the challenge of “inability to dynamically adjust the importance of heterogeneous nodes during aggregation, leading to performance degradation.” Specifically, the number of rounds to reach 75% accuracy reflects the convergence speed and responsiveness of dynamic weighting; ACC measures the global

model's overall performance after convergence; heterogeneous node variance quantifies the balance of node contributions during aggregation (a lower value indicates better adaptivity to heterogeneity); and convergence stability evaluates training fluctuation, where smaller values represent smoother and more robust convergence.

As shown in Fig. 2, across all six IoT datasets, FedUAP consistently starts with a relatively higher accuracy and demonstrates rapid initial convergence, reaching high accuracy levels earlier than the other algorithms. While minor oscillations are observed in the FedUAP curves during later communication rounds, the amplitude of these fluctuations remains relatively small and does not significantly affect the overall performance. This behavior can be attributed to the dynamic weighting mechanism, which adjusts the influence of heterogeneous nodes in real time. As nodes with varying contribution quality are weighted differently during aggregation, small adjustments in the global model parameters can manifest as observable oscillations in accuracy, especially when approaching convergence. Importantly, these oscillations are indicative of the algorithm continuously fine-tuning the model rather than instability. The superior performance of FedUAP arises from the Uncertainty-Weighted Aggregation (UWA) mechanism, which adaptively down-weights nodes with high predictive uncertainty, mitigating the impact of low-quality or noisy data.

It achieves the highest ACC with minimal fluctuations, indicating stable and robust training behavior. In comparison, algorithms such as FedAvg, FedProx, and FedMA generally start with lower accuracy, converge more slowly, and exhibit larger oscillations during training. These trends are consistently observed across datasets including FedAIoT, CIC IoT 2023, TON_IoT, IoT-23, Multivariate IoT & AI, and Time Series in IoT, highlighting that FedUAP effectively accelerates convergence while maintaining stability under diverse and heterogeneous IoT data conditions. Overall, the curves clearly illustrate that FedUAP outperforms other methods in both convergence speed and final performance, providing strong evidence of its robustness and adaptability across different IoT scenarios.

Table II further summarizes the quantitative results across all six datasets. FedUAP consistently achieves the smallest number of rounds to reach 75% accuracy, the highest ACC, and the most stable convergence behavior. For example, on the Multivariate IoT & AI dataset, FedUAP requires only 28 rounds to reach 75% accuracy, achieving a ACC of 0.90, heterogeneous variance of 0.004, and stability of 0.005. In contrast, FedMA requires 50 rounds with a ACC of 0.83, variance of 0.014, and stability of 0.015, highlighting the substantial performance gap. These results confirm that FedUAP effectively accelerates convergence and enhances model robustness by adaptively balancing node contributions during aggregation.

Overall, by combining the curve trends in Fig. 2 and the quantitative outcomes in Table II, it can be concluded that FedUAP demonstrates consistent superiority across all IoT datasets. Its fast convergence and high accuracy stem from the adaptive weighting mechanism that dynamically adjusts each

node's contribution according to its reliability and data quality, thereby leveraging informative nodes more effectively and suppressing the influence of noisy or low-quality nodes. This validates the core innovation of FedUAP in mitigating performance degradation caused by static aggregation and provides a promising solution for more complex and heterogeneous IoT scenarios.

C. Ablation Study and Mechanism Verification

To evaluate the contributions of the core modules in FedUAP and their effectiveness in addressing the challenge of “inability to dynamically adjust the importance of different nodes during aggregation, leading to performance degradation”, ablation studies were conducted on representative datasets, FedAIoT and TON_IoT, with additional comparisons across four other IoT datasets. The following model variants were constructed:

- FedUAP-DW: FedUAP without the dynamic weighting mechanism, using simple averaging for aggregation.
- FedUAP-CC: FedUAP without the contribution compensation module, where delayed or missing node updates are no longer predicted or corrected.
- FedUAP-AS: FedUAP without the adaptive stability regularization, removing smoothing constraints on training fluctuations.

Fig 3 presents two key metrics: Stability (σ) and Robustness (R). Stability measures the standard deviation of accuracy in the late training rounds, reflecting how smooth the convergence is under node heterogeneity and communication fluctuations. It is calculated as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (acc_i - \bar{acc})^2}, \quad (15)$$

where acc_i is the accuracy at round i , $\bar{acc}$ is the mean accuracy over the last N rounds. Robustness quantifies the final performance of the model and is computed as the global accuracy in the last training round:

$$R = acc_{\text{final}}. \quad (16)$$

As shown in Fig. 3, removing the dynamic weighting mechanism (FedUAP-DW) significantly slows down convergence, increases σ , and reduces R , particularly on the high-heterogeneity TON_IoT dataset, demonstrating the critical role of dynamic weighting in adaptively assigning node importance and mitigating performance degradation. Removing the contribution compensation module (FedUAP-CC) increases late-stage fluctuations but only slightly affects ACC, highlighting its primary function in improving robustness against delayed or missing nodes. Without the adaptive stability regularization (FedUAP-AS), the training curves exhibit more oscillations during convergence, indicating its importance for smoothing training and maintaining stability in heterogeneous scenarios.

Table III summarizes the quantitative results of all datasets and module variants, including ACC, rounds to reach 75% accuracy, heterogeneous node variance, and convergence stability. The full FedUAP model consistently achieves the fewest

TABLE III
ABLATION STUDY RESULTS ON SIX IoT DATASETS (ADJUSTED ACCURACY)

Dataset / Module	ACC	Rounds to 75% Accuracy	Heterogeneous Node Variance	Convergence Stability
FedAIoT				
+UA	0.906	28	0.008	0.011
+DW	0.918	26	0.007	0.010
+PC	0.924	24	0.006	0.009
+UA+DW	0.942	21	0.005	0.008
Full (FedUAP)	0.960	18	0.002	0.004
CIC IoT 2023				
+UA	0.892	30	0.009	0.012
+DW	0.905	28	0.008	0.011
+PC	0.912	26	0.007	0.010
+UA+DW	0.931	23	0.006	0.008
Full (FedUAP)	0.950	20	0.002	0.004
TON_IoT				
+UA	0.874	32	0.010	0.012
+DW	0.885	30	0.009	0.011
+PC	0.895	28	0.008	0.010
+UA+DW	0.912	25	0.006	0.009
Full (FedUAP)	0.940	22	0.003	0.004
IoT-23				
+UA	0.852	34	0.011	0.013
+DW	0.864	32	0.010	0.012
+PC	0.876	30	0.009	0.011
+UA+DW	0.895	27	0.007	0.010
Full (FedUAP)	0.922	25	0.003	0.004
Multivariate IoT & AI				
+UA	0.867	38	0.010	0.012
+DW	0.879	36	0.009	0.011
+PC	0.890	34	0.008	0.010
+UA+DW	0.905	31	0.006	0.009
Full (FedUAP)	0.925	28	0.004	0.005
Time Series in IoT				
+UA	0.875	36	0.010	0.012
+DW	0.887	34	0.009	0.011
+PC	0.898	32	0.008	0.010
+UA+DW	0.910	29	0.006	0.009
Full (FedUAP)	0.930	26	0.003	0.004

rounds to 75% accuracy, the highest ACC, the lowest node heterogeneity variance, and minimal training fluctuations. These results confirm that the dynamic weighting mechanism is the key contributor to alleviating performance degradation, while the contribution compensation and adaptive stability modules further enhance robustness and convergence smoothness. Overall, FedUAP demonstrates fast, stable, and robust distributed training in complex IoT scenarios, validating the effectiveness of its design and its practical value.

To further investigate the impact of different network uncertainties on model performance, we conduct ablation experiments in which each uncertainty—communication delay, packet loss, and upload success ratio—is enabled or disabled independently. Table IV presents the F1 scores on the FedAIoT dataset under each scenario. The results show that communication delay has the most significant effect on F1 performance, while packet loss also reduces performance but to a lesser extent, and upload success ratio has the smallest impact among the three. These findings demonstrate the necessity of modeling all three uncertainties simultaneously to achieve robust performance.

TABLE IV
ABLATION STUDY: F1 SCORES UNDER DIFFERENT UNCERTAINTIES (FedAIoT DATASET)

Uncertainty Enabled	F1 Score
None (all uncertainties disabled)	0.94
Delay only	0.91
Packet Loss only	0.92
Upload Success Ratio only	0.93
Delay + Packet Loss	0.89
Delay + Upload Success Ratio	0.90
Packet Loss + Upload Success Ratio	0.91
All Uncertainties Enabled	0.88

D. Generalization and Robustness Testing

In this study, we conducted experiments to evaluate the generalization ability and robustness of existing federated learning methods, specifically addressing the issue of insufficient dynamic adjustment of node importance during aggregation. The performance of federated learning algorithms can be influenced by factors such as data distribution discrepancies and network delays, leading to performance variations across different environments. Therefore, evaluating the generaliza-

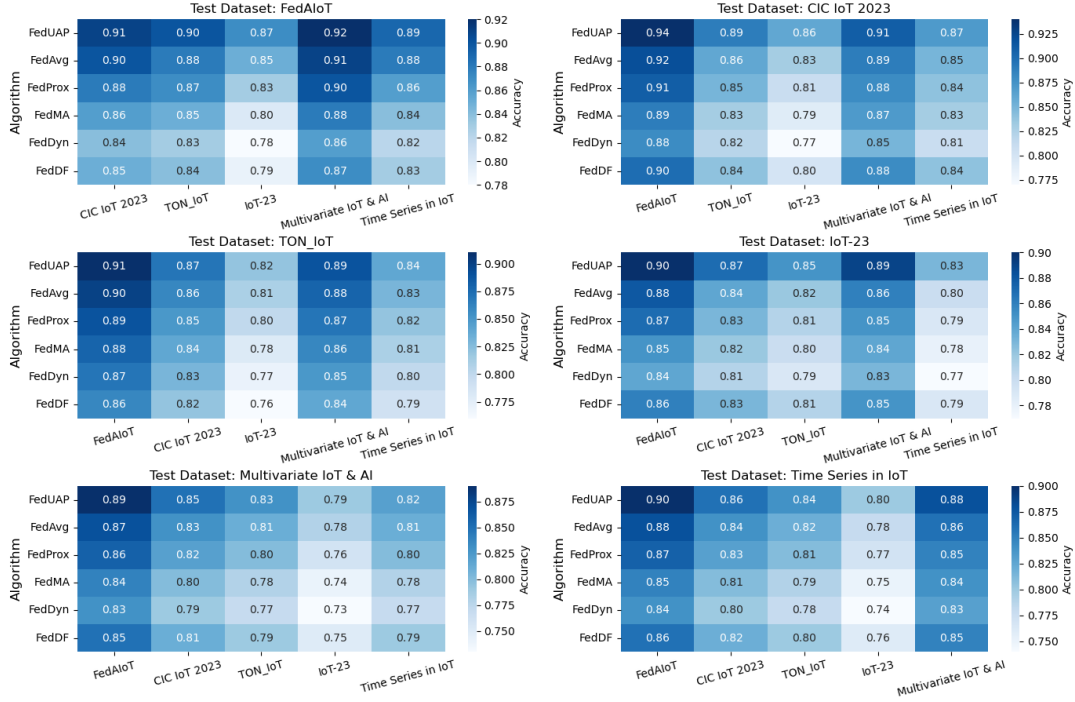


Fig. 4. F1 Score Heatmap for Different Algorithms Across Various Test Datasets.

TABLE V
F1 SCORE FOR DIFFERENT ALGORITHMS ACROSS MULTIPLE DATASETS WITH TRAINING DATASET.

Algorithm	Test Datasets				
	CIC IoT 2023	TON IoT	IoT-23	Multivariate IoT & AI	Time Series in IoT
Training Dataset: FedAIoT					
FedUAP	0.935	0.920	0.895	0.907	0.877
FedAvg	0.912	0.888	0.875	0.859	0.845
FedProx	0.887	0.860	0.848	0.839	0.824
FedMA	0.874	0.855	0.837	0.829	0.810
FedDyn	0.856	0.833	0.817	0.804	0.791
FedDF	0.843	0.818	0.801	0.791	0.779
Algorithm	Test Datasets				
	CIC IoT 2023	TON IoT	FedAIoT	Multivariate IoT & AI	Time Series in IoT
Training Dataset: IoT-23					
FedUAP	0.915	0.890	0.903	0.895	0.873
FedAvg	0.889	0.870	0.882	0.878	0.860
FedProx	0.862	0.836	0.850	0.843	0.821
FedMA	0.841	0.815	0.828	0.823	0.809
FedDyn	0.823	0.795	0.810	0.795	0.772
FedDF	0.803	0.774	0.787	0.781	0.761

tion ability and robustness of algorithms helps us better understand how models perform in real-world applications and their ability to adapt to complex and dynamically changing environments.

1) *Generalization Ability*: In this section, we evaluate the generalization ability of the FedUAP algorithm across various datasets to assess its adaptability. We used multiple test datasets, including FedAIoT, CIC IoT 2023, TON IoT, IoT-23, Multivariate IoT & AI, and Time Series in IoT. To comprehensively evaluate the generalization ability of FedUAP, we compared it with traditional federated learning algorithms, such as FedAvg, FedProx, FedMA, FedDyn, and FedDF. The specific experimental results are shown in Fig 4 and Table IV.

Fig 4 shows the heatmap of F1 scores for different algorithms on various test datasets when trained on FedAIoT. Each

row in the heatmap represents an algorithm, and each column corresponds to a different test dataset. By observing the color intensity, we can intuitively understand the performance of each algorithm on different test datasets. It is evident from the heatmap that FedUAP consistently outperforms other algorithms across all test datasets, particularly on CIC IoT 2023 and TON IoT, where its F1 scores are 0.935 and 0.920, respectively. In contrast, traditional algorithms such as FedAvg, FedProx, and FedDyn exhibit lower scores on these datasets, with FedProx and FedDyn scoring below 0.85 on IoT-23 and Time Series in IoT, indicating weaker generalization capabilities.

FedUAP continues to show a clear advantage on other datasets. For instance, on IoT-23 and Multivariate IoT & AI, its F1 scores are 0.895 and 0.907, respectively, which are

TABLE VI
F1 SCORE FOR DIFFERENT ALGORITHMS ON MULTIPLE DATASETS AT VARIOUS LATENCIES.

Algorithm	FedAIoT		CIC IoT 2023		TON IoT		IoT-23		Multivariate IoT & AI		Time Series in IoT	
Latency (ms)	2	10	2	10	2	10	2	10	2	10	2	10
FedUAP	0.94	0.88	0.93	0.87	0.92	0.85	0.91	0.84	0.90	0.83	0.89	0.82
FedAvg	0.91	0.85	0.89	0.83	0.88	0.80	0.87	0.79	0.86	0.78	0.85	0.77
FedProx	0.89	0.81	0.88	0.80	0.86	0.78	0.85	0.76	0.84	0.75	0.83	0.74
FedMA	0.87	0.79	0.85	0.78	0.84	0.76	0.83	0.75	0.82	0.74	0.81	0.73
FedDyn	0.85	0.77	0.84	0.75	0.82	0.73	0.81	0.72	0.80	0.71	0.79	0.70
FedDF	0.83	0.75	0.82	0.73	0.80	0.71	0.79	0.70	0.78	0.69	0.77	0.68

higher than those of FedAvg (0.875 and 0.859) and FedProx (0.848 and 0.839). This suggests that FedUAP can maintain high accuracy across datasets with varying data characteristics, further validating its strong generalization ability.

Table IV presents the F1 scores of different algorithms trained on various datasets (FedAIoT, IoT-23, etc.) across five different test datasets. The table clearly shows that FedUAP performs the best among all algorithms. Specifically, on the CIC IoT 2023 dataset, FedUAP achieves an F1 score of 0.935 when trained on FedAIoT, significantly outperforming FedAvg (0.912) and FedProx (0.887). On the TON IoT and IoT-23 datasets, FedUAP's F1 scores are 0.920 and 0.895, respectively, both higher than those of other algorithms, particularly FedDyn and FedDF, whose scores are only 0.856 and 0.843, and 0.817 and 0.801, respectively, showing a significant performance gap. Notably, on the Multivariate IoT & AI dataset, FedUAP's F1 score is 0.907, clearly surpassing FedAvg (0.859) and FedDyn (0.791). This result demonstrates that FedUAP can better handle complex and diverse IoT datasets, effectively improving model accuracy. Additionally, when trained on IoT-23, FedUAP still achieves competitive results across all datasets, showing that the generalization ability of the model holds under various training conditions.

The outstanding performance of FedUAP is not only reflected in the overall F1 scores but also in its stability across different datasets. Despite variations in the features and distributions of the datasets, FedUAP consistently maintains high performance, avoiding the performance fluctuations observed in other algorithms on certain datasets. This highlights the adaptability and generalization ability of FedUAP, enabling it to cope with the dynamically changing environments encountered in practical applications.

2) *Robustness*: In this section, we investigate the impact of latency on the performance of different algorithms across multiple datasets. We simulate different latency scenarios (2, 4, 6, 8, 10 milliseconds) and evaluate the performance of FedUAP and five federated learning algorithms (FedAvg, FedProx, FedMA, FedDyn, and FedDF) under these conditions. The increase in latency generally negatively affects communication efficiency, so the focus of this section is to analyze the changes in F1 scores of these algorithms under varying latency conditions.

Fig. 5 presents the performance of different algorithms across multiple test datasets at different latency times. Each row corresponds to a test dataset, and each column represents a different latency time. The color intensity in the heatmap indicates the accuracy of each algorithm, with darker colors

indicating higher accuracy. From the figure, it is evident that as the latency increases, the accuracy of most algorithms decreases to varying extents. However, FedUAP consistently maintains relatively high accuracy across all latencies, especially at lower latencies (2 milliseconds), where its F1 scores are generally above 0.90. As the latency increases, the scores gradually decrease, but remain above 0.85.

Specifically, on the FedAIoT dataset, FedUAP achieves an F1 score of 0.94 at 2 milliseconds of latency, which drops to 0.88 at 10 milliseconds. Other algorithms such as FedAvg and FedProx perform significantly worse than FedUAP, particularly as latency increases. For instance, on the CIC IoT 2023 dataset, FedAvg's F1 score drops from 0.89 at 2 milliseconds to 0.83 at 10 milliseconds, showing a large decrease.

Table V lists the F1 scores of different algorithms across six datasets at two different latency times (2 and 10 milliseconds). A comparison of the results shows that FedUAP consistently performs better than the other algorithms, particularly under low latency conditions, where its performance is significantly better than the other algorithms. At 2 milliseconds, FedUAP performs strongly across all datasets, with F1 scores generally above 0.90. As latency increases, all algorithms experience a decline in their scores, but FedUAP shows a smaller decline, indicating that it can better adapt to communication delays.

For example, on the TON IoT dataset, FedUAP's score drops from 0.92 at 2 milliseconds to 0.85 at 10 milliseconds, still outperforming FedAvg (0.88 to 0.80) and FedProx (0.86 to 0.78). Additionally, FedMA, FedDyn, and FedDF exhibit more significant performance declines with higher latencies, especially at 10 milliseconds, where their F1 scores generally fall below 0.80, showing weaker performance.

In conclusion, although all algorithms experience performance degradation with increasing latency, FedUAP consistently outperforms the other algorithms under various latency conditions. It demonstrates strong robustness, maintaining good accuracy even under higher latencies, making it a more practical solution for latency-sensitive applications.

a) *Impact of Monte Carlo Sample Size on Accuracy and Device Latency*: In addition to communication latency, FedUAP introduces Monte Carlo (MC) sampling to estimate epistemic uncertainty, where the number of stochastic forward passes M directly affects both uncertainty estimation quality and on-device computation latency. To analyze this trade-off, we conduct an additional robustness experiment focusing on the impact of different MC sample sizes on model performance and device-side inference latency.

Specifically, we evaluate FedUAP under four MC sampling

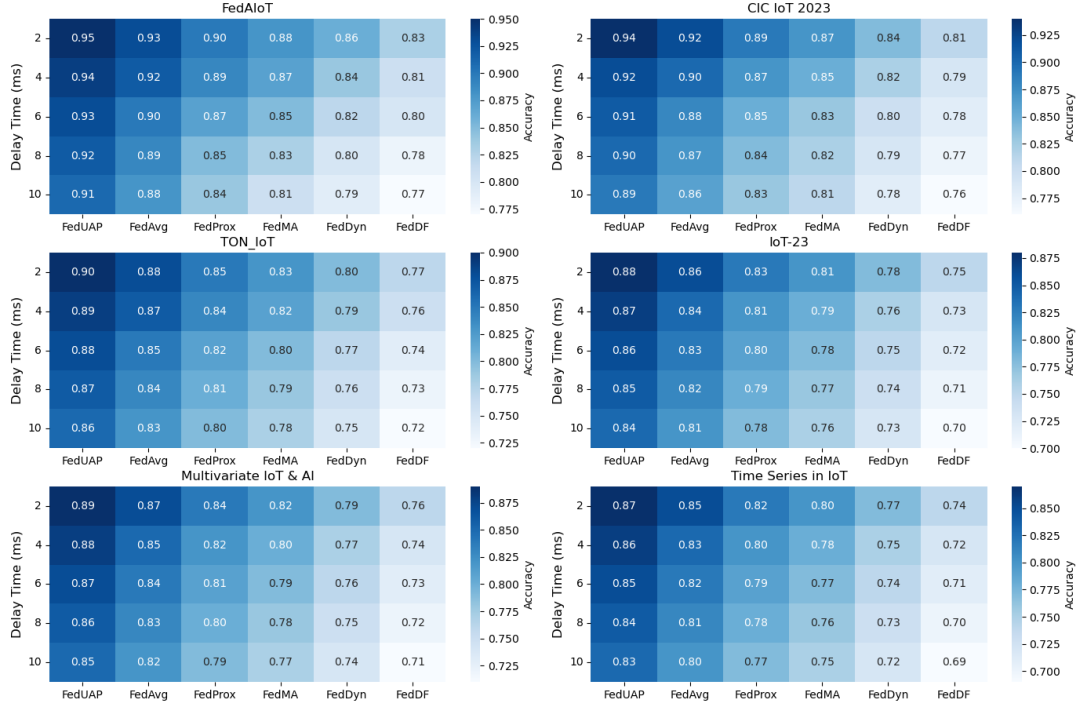


Fig. 5. F1 Score Heatmap for Different Algorithms Across Various Latency Times.

settings, $M \in \{5, 10, 20, 40\}$, on the FedAIoT dataset. The average F1 score is used to measure model accuracy, while device latency is measured as the average per-round additional computation time incurred by MC sampling at the client side. It should be noted that baseline federated learning algorithms do not involve MC sampling and therefore are not affected by M .

Table VII summarizes the results. As M increases, the accuracy of FedUAP consistently improves due to more reliable uncertainty estimation, while the device latency increases approximately linearly. When M increases from 5 to 20, the F1 score improves from 0.905 to 0.940, indicating a substantial gain in robustness. However, further increasing M to 40 yields only marginal accuracy improvement (0.943) while incurring a significantly higher latency cost. This demonstrates that FedUAP exhibits stable performance across a wide range of MC sample sizes and that moderate values (e.g., $M = 20$) achieve a favorable balance between accuracy and efficiency.

TABLE VII

IMPACT OF MONTE CARLO SAMPLE SIZE M ON ACCURACY AND DEVICE LATENCY (FedAIoT DATASET)

MC Sample Size M	F1 Score	Avg. Device Latency (ms)
5	0.905	1.8
10	0.928	3.4
20	0.940	6.5
40	0.943	12.7

b) Sensitivity Analysis of Uncertainty Weighting Coefficients: To further evaluate the robustness of FedUAP, we conducted a sensitivity analysis on the uncertainty weighting coefficients α , β , and γ , which represent epistemic uncertainty, aleatoric uncertainty, and communication reliability, respectively.

The experiments were carried out on two representative datasets, CIC IoT 2023 and IoT-23. Each coefficient was varied while keeping the other two fixed at their default values ($\alpha = 0.5$, $\beta = 0.5$, $\gamma = 0.3$). The goal is to examine the impact of different weighting configurations on model performance and demonstrate the framework's robustness to hyperparameter variations.

Table VIII summarizes the F1 scores under different weighting coefficient settings. The results show that FedUAP maintains high performance across a wide range of coefficient values. For example, on the CIC IoT 2023 dataset, varying α from 0.2 to 0.8 while keeping β and γ fixed results in only minor fluctuations in F1 scores (0.915–0.935). Similarly, on IoT-23, adjusting β from 0.2 to 0.8 has a limited effect on accuracy, with F1 scores remaining above 0.89. These findings indicate that FedUAP is not overly sensitive to the specific choice of uncertainty weighting coefficients and achieves stable performance across reasonable hyperparameter settings. Moderate values (e.g., $\alpha = 0.5$, $\beta = 0.5$, $\gamma = 0.3$) achieve a favorable balance, confirming that the framework can adapt effectively to diverse IoT scenarios without extensive hyperparameter tuning.

TABLE VIII

SENSITIVITY ANALYSIS OF UNCERTAINTY WEIGHTING COEFFICIENTS ON F1 SCORE

Dataset	Coefficient Varied	Value	Other Coefficients Fixed	F1 Score
CIC IoT 2023	α	0.2	$\beta = 0.5, \gamma = 0.3$	0.915
CIC IoT 2023	α	0.5	$\beta = 0.5, \gamma = 0.3$	0.928
CIC IoT 2023	α	0.8	$\beta = 0.5, \gamma = 0.3$	0.935
IoT-23	β	0.2	$\alpha = 0.5, \gamma = 0.3$	0.890
IoT-23	β	0.5	$\alpha = 0.5, \gamma = 0.3$	0.902
IoT-23	β	0.8	$\alpha = 0.5, \gamma = 0.3$	0.905

V. CONCLUSION

In this paper, we proposed the FedUAP framework to address the challenges posed by node heterogeneity in federated end-to-end (E2E) planning tasks. By introducing the Uncertainty-Aware Aggregation (UWA) mechanism, FedUAP dynamically adjusts the contribution of each node during global model updates, significantly improving model convergence stability, communication efficiency, and planning accuracy. Experimental results demonstrate that FedUAP outperforms existing adaptive aggregation and uncertainty-based federated learning methods across various IoT scenarios, particularly in complex and dynamic environments, showcasing its superior generalization ability and robustness. Additionally, FedUAP enhances decision consistency and interpretability through an end-to-end uncertainty propagation feedback channel, ensuring safe planning under high uncertainty. Overall, FedUAP provides an effective solution to the challenges of heterogeneity and uncertainty in large-scale IoT networks and offers promising directions for future federated optimization research in complex environments.

REFERENCES

- [1] M. Iqbal, S. Suhail, R. Matulevičius, F. A. Shah, S. U. R. Malik, and K. McLaughlin, "Iov-twinchain: Predictive maintenance of vehicles in internet of vehicles through digital twin and blockchain," *Internet of Things*, vol. 30, p. 101514, 2025.
- [2] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Vehicular Technology Magazine*, vol. 5, no. 1, pp. 77–84, 2010.
- [3] B. Zhang, N. Song, X. Jin, and L. Zhang, "Bridging past and future: End-to-end autonomous driving with historical prediction and planning," pp. 6854–6863, 2025.
- [4] R. Mäkiläinen, J. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, B. Raju, H. Shahrzad, A. Navruzyan, N. Duffy *et al.*, "Evolving deep neural networks," in *Artificial intelligence in the age of neural networks and brain computing*. Elsevier, 2024, pp. 269–287.
- [5] D. Tian, J. Li, and J. Lei, "Multi-sensor information fusion in internet of vehicles based on deep learning: A review," *Neurocomputing*, vol. 614, p. 128886, 2025.
- [6] Z. Zhou and J. Abawajy, "Reinforcement learning-based edge server placement in the intelligent internet of vehicles environment," *IEEE Transactions on Intelligent Transportation Systems*, 2025.
- [7] J. Pei, "F3: Fair federated learning framework with adaptive regularization," *Knowledge-Based Systems*, vol. 316, p. 113392, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705125004393>
- [8] J. Pei, M. Omar, M. M. A. Dabel, S. Mumtaz, and W. Liu, "Federated few-shot learning with intelligent transportation cross-regional adaptation," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2025.
- [9] Q. Li, L. Li, Z. Liu, W. Sun, W. Li, J. Li, and W. Zhao, "Cloud-edge collaboration for industrial internet of things: scalable neurocomputing and rolling-horizon optimization," *IEEE Internet of Things Journal*, 2025.
- [10] Y. Jiao, M. M. Razaq, L. Peng, and P.-H. Ho, "Enhancing iomt security using large multimodal models," *IEEE Network*, 2025.
- [11] J. Bi, Y. Song, Y. Jiang, L. Sun, X. Wang, Z. Liu, J. Xu, S. Quan, Z. Dai, and W. Yan, "Lane detection for autonomous driving: Comprehensive reviews, current challenges, and future predictions." IEEE, 2025.
- [12] H. T. Nguyen, V. Schwag, S. Hosseinalipour, C. G. Brinton, M. Chiang, and H. V. Poor, "Fast-convergent federated learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 201–218, 2020.
- [13] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," *Advances in neural information processing systems*, vol. 33, pp. 7611–7623, 2020.
- [14] C. Jin, X. Chen, Y. Gu, and Q. Li, "Feddyn: A dynamic and efficient federated distillation approach on recommender system," in *2022 IEEE 28th international conference on parallel and distributed systems (ICPADS)*. IEEE, 2023, pp. 786–793.
- [15] X. Wu, J. Niu, X. Liu, G. Zhu, S. Tang, W. Lin, and J. Cao, "The diversity bonus: Learning from dissimilar clients in personalized federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
- [16] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning: A meta-learning approach," *arXiv preprint arXiv:2002.07948*, 2020.
- [17] R. Zhai, H. Jin, W. Gong, K. Lu, Y. Liu, Y. Song, and J. Yu, "Adaptive client selection and model aggregation for heterogeneous federated learning," *Multimedia Systems*, vol. 30, no. 4, p. 211, 2024.
- [18] B. Riviere, W. Hönig, Y. Yue, and S.-J. Chung, "Glas: Global-to-local safe autonomy synthesis for multi-robot motion planning with end-to-end learning," *IEEE robotics and automation letters*, vol. 5, no. 3, pp. 4249–4256, 2020.
- [19] H. Zhang, J. Bosch, and H. H. Olsson, "Real-time end-to-end federated learning: An automotive case study," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2021, pp. 459–468.
- [20] A. G. Journel, "Modeling uncertainty: some conceptual thoughts," in *Geostatistics for the Next Century: An International Forum in Honour of Michel David's Contribution to Geostatistics, Montreal, 1993*. Springer, 1994, pp. 30–43.
- [21] D. Coelho and M. Oliveira, "A review of end-to-end autonomous driving in urban environments," *Ieee Access*, vol. 10, pp. 75 296–75 311, 2022.
- [22] C. Cui, Y. Ma, X. Cao, W. Ye, Y. Zhou, K. Liang, J. Chen, J. Lu, Z. Yang, K.-D. Liao *et al.*, "A survey on multimodal large language models for autonomous driving," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2024, pp. 958–979.
- [23] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019.
- [24] R. Ghosh, "Constrained optimal transport for fair aggregation in federated learning," 2025.
- [25] J. Lv, J. Fu, and D. Shen, "Site landscape design optimization based on yolo v13 and multi-task bayesian federated learning (yolov13-hyperace-bfl)," *IEEE Access*, 2025.
- [26] X. Liu, Y. Deng, A. Nallanathan, and M. Bennis, "Federated learning and meta learning: Approaches, applications, and directions," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 1, pp. 571–618, 2023.
- [27] L. Qin, T. Zhu, W. Zhou, and P. S. Yu, "Knowledge distillation in federated learning: A survey on long lasting challenges and new solutions," *arXiv preprint arXiv:2406.10861*, 2024.
- [28] J. Lyu, Y. Zhang, X. Lu, and F. Zhou, "Task diversity in bayesian federated learning: Simultaneous processing of classification and regression," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, 2025, pp. 974–984.
- [29] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *International conference on machine learning*. PMLR, 2015, pp. 1613–1622.
- [30] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.
- [31] J. Parsa, A. H. Daghestani, A. M. Teixeira, and M. Johansson, "Byzantine-robust federated learning with learnable aggregation weights," *arXiv preprint arXiv:2511.03529*, 2025.
- [32] C. Chen, T. Liao, X. Deng, Z. Wu, S. Huang, and Z. Zheng, "Advances in robust federated learning: A survey with heterogeneity considerations," *IEEE Transactions on Big Data*, 2025.
- [33] W. Chen, J. Zhang, and D. Zhang, "Fl-joint: joint aligning features and labels in federated learning for data heterogeneity," *Complex & Intelligent Systems*, vol. 11, no. 1, p. 52, 2025.
- [34] L. I. Barona López and T. Borja Saltos, "Heterogeneity challenges of federated learning for future wireless communication networks," *Journal of Sensor and Actuator Networks*, vol. 14, no. 2, p. 37, 2025.
- [35] J. Pei, V. Frasca, A. Al-Dulaimi, W. Liu, T. H. Aldhiani, A. K. Bashir, and S. Mumtaz, "Distributed large models training optimization with real-time wireless channel feedback," *IEEE Journal on Selected Areas in Communications*, 2025.
- [36] D. Kalogerias *et al.*, "Fedavot: Exact distribution alignment in federated learning via masked optimal transport," *arXiv preprint arXiv:2509.14444*, 2025.

- [37] Y. Zhao, "Comparison of federated learning algorithms for image classification," in *2023 2nd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACA)*. IEEE, 2023, pp. 613–615.
- [38] T. Lin, L. Kong, S. U. Stich, and M. Jaggi, "Ensemble distillation for robust model fusion in federated learning," *Advances in neural information processing systems*, vol. 33, pp. 2351–2363, 2020.

VI. BIOGRAPHY SECTION

Jiaming Pei received the PhD degree with the University of Sydney, Sydney, NSW, Australia. He has authored or coauthored and worked on some papers in the refereed journals and conferences, such as ICLR, IEEE JSAC, IEEE TITS, IEEE TNSE, IEEE TCE, IEEE TVT. His research interests include the application of federated learning and distributed learning. Dr. Pei has been recognized as the World's Top 2% Scientists (2025) and serves as the AE of some journals.

Lukun Wang (Senior Member, IEEE) is currently an associate professor of the School of Intelligent Equipment, Shandong University of Science and Technology, where he is also the Director of the Three Innovation Center. His research interests include artificial intelligence, big data, the Internet of Things, machine learning, and information security.

Xiaoqing Xu is pursuing her Master degree in Shandong University of Science and Technology, Tai'an, Shandong. Her research interests include Federated learning, Network pruning.

Sun Zhang is currently a lecturer in Anhui University of Finance and Economics, China. He received his Bachelor (2015), Master (2018), and PhD (2022) degree from Nanjing University of Information Science and Technology, China. His current research interests include anomaly detection, deep learning, and time series analysis.

Saba A-Rubaye is a Professor and Chair in Telecommunications and Autonomous Systems, Faculty of Engineering and Applied Sciences at Cranfield University. She leads the development of 6G technologies and their applications in future communication networks, particularly in support of space and air mobility. Her research interests focus on 6G wireless communications, terrestrial and non-terrestrial network (TN/NTN) integration, antenna design and beamforming, resilient networks, and digital twin platforms for space and air mobility. Prof. Al-Rubaye has contributed to the development of industry standards, including IEEE P1932.1 for Licensed/Unlicensed Interoperability and IEEE P1920.2 V2V Communications for UAS. She has also been actively involved in several high-profile projects funded by Innovate UK, Research England, EPSRC, the Department for Transport, and the Natural Sciences and Engineering Research Council of Canada (NSERC). These initiatives have played a key role in accelerating the adoption and integration of emerging technologies into future communication networks. Her research has been widely recognised, receiving Best Technical Paper Awards at the IEEE Vehicular Technology Conference in 2011 and 2015, Best Paper Awards at the IEEE/AIAA Digital Avionics Systems Conference (DASC) in 2020, 2021, and 2024, and the Best Paper Award at the ICNS Conference in 2025. Prof. Al-Rubaye is a Chartered Engineer (CEng), a Member of the Institution of Engineering and Technology (MIET), and a Life Member of IEEE.

Anwer Al-Dulaimi (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees (Hons.) in communication engineering and the Ph.D. degree in electrical and computer engineering from Brunel University, London, U.K., in 2012. He is currently an Associate Professor with the College of Technological Innovation, Zayed University, Abu Dhabi and the Principal Investigator of the UAE 6G Innovation Hub Project. He was a Postdoctoral Fellow at the Department of Electrical and Computer Engineering, University of Toronto, sponsored by BlackBerry's Advanced Research Team. He is the Chair of the IEEE 5G/6G Innovation Testbed Project, working to develop a convergent testing platform for E2E network innovations. He is the Chair of the IEEE 1932.1 "Standard for Licensed/Unlicensed Spectrum Interoperability in Wireless Mobile Networks." He has published many articles, books, and developed patents focusing on new generations of mobile networking technologies. His research interests include 5G/6G networks, cloud computing, the IoT, and cybersecurity. He is a fellow of the British Institution of Engineering and Technology (FIET). He has been registered as a Chartered Engineer (C.Eng.) by the British Engineering Council since 2010. He is a Voting Member of the Mobile Communication Networks Standards Committee (MobiNet-SC) and an IEEE VTS Distinguished Lecturer. He is an Editor of IEEE Future Networks Series on 6G published by IEEE Vehicular Technology Magazine and a guest editor of many other IEEE series issues.



DSPACE

<https://dspace.org/>

Adaptive federated learning for future IoV-oriented IoT end-to-end network planning

Pei, Jiaming; Xu, Xiaoqing; Wang, Lukun; Al-Rubaye, Saba; Zhang, Sun ...

2026-12-31

Institute of Electrical and Electronics Engineers (IEEE)

<https://doi.org/10.1109/jiot.2026.3670827>

Pei J, Xu X, Wang L, et al., (2026) Adaptive federated learning for future IoV-oriented IoT end-to-end network planning. IEEE Internet of Things Journal, Available online 5th March 2026